Lecture notes for 6CCM338A
Mathematical Finance II: Continuous Time

Leandro Sánchez-Betancourt *

March 14, 2023

*Based on the lecture notes from previous years by Dr. Martin Forde. Some of the material on applied probability is taken from lectures notes by Prof. Markus Riedle. My email address is leandro.sanchez-betancourt@kcl.ac.uk. The GitHub repository for the course is at `https://github.com/leandro-sbetancourt/kcl-continuous-time-finance`. In there you will be able to find the code that produces most figures and examples in the lecture notes.

Contents

1 Introduction **1**

2 Probability essentials **4**
2.1 Kolmogorov's axioms 4
2.2 Random variables, expectation, and other useful concepts 6
2.3 Stochastic processes 13
2.4 Conditioning 15
2.5 Other interesting concepts 16
2.6 Example questions 17

3 Brownian motion **19**
3.1 Elementary properties of Brownian motion 19
3.2 Lack of differentiability 22
3.3 Constructing and simulating Brownian motion 22
3.4 Quadratic variation of Brownian motion 25

4 Stochastic integrals with respect to Brownian motion **29**
4.1 Why do we need a new type of integration? 29
4.2 The construction 30
4.3 Stochastic differential equations 32
4.4 Ito's lemma 33
4.5 Important examples 34

5 Continuous-time models **38**
5.1 The Black-Scholes model 38
5.2 The terminal share price distribution 38
5.3 The Black-Scholes PDE 40
5.4 The Black-Scholes formula for a European Call option 43
5.4.1 Numerical example 44
5.5 The Greeks 45
5.6 The Feynman-Kac formula 45
5.6.1 Implied volatility 47
5.7 Continuous-time martingales 48

6 Pricing with simulations **50**
6.1 Computing integrals using simulations 50
6.2 Pricing derivatives 52
6.3 Pricing a path dependent derivative 53

7 Barrier options (not examinable) **56**
7.1 Semi-static hedging: pricing and hedging barrier options 56
7.2 PDE approach: pricing and hedging barrier options 59
7.3 Simulations and the advantage of the the semi-static argument 59

References **61**

8 Appendix **62**
8.1 A note on continuous compounding 62
8.2 The self-financing equation 64
8.3 Calculations for the semi-static hedging argument 65

1 Introduction

Welcome to 6CCM338. This course provides an introduction to mathematical finance for continuous time models. Continuous time models are both challenging and interesting, and the financial models and instruments you will learn about are used on real life trading desks, where trillions of dollars worth of derivatives trade hands every day. Many previous students of the course now work in the city of London; others have gone to do the MSc in Financial Mathematics here or elsewhere and eventually a PhD. This course will give you many tools that will prove useful if you go to the industry or pursue more advanced studies, but the course is not easy.

In this course, you will learn about:

(i) **Brownian motion**, a continuous time random process W_t which has independent, identically distributed increments with the property that $W_t - W_s \sim N(0, t-s)$ for $s < t$, i.e. Normally distributed with mean zero and variance $t - s$. Brownian motion has a continuous sample paths (i.e. it does not jump) but is not differentiable with respect to t, so we cannot apply standard calculus techniques to integrate with respect to W_t, so we have to develop a new calculus called **stochastic calculus**.

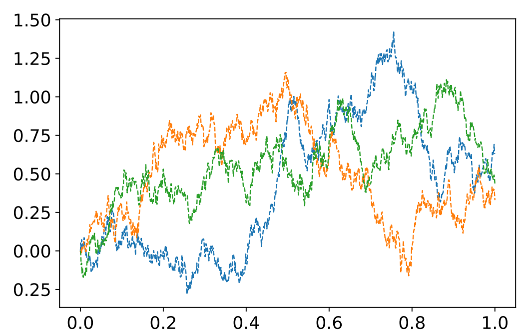

Figure 1: Three simulations of a standard Brownian motion.

(ii) **Ito's lemma** - how to do calculus on $f(W_t)$, where f is a twice differentiable function- we will see the rules are subtly different to standard 1st year calculus.

(iii) The **Black-Scholes model** for a share price process S_t:

$$S_t = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t} \tag{1}$$

where W_t is Brownian motion, μ is the drift of the process, which describes the overall upward/downward trend, and σ is the volatility, which describes its variability.

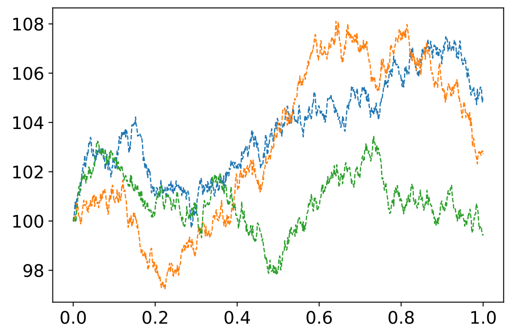

Figure 2: Three simulations of a geometric Brownian motion with $S_0 = 100$, $\mu = 0.05$, and $\sigma = 0.1$.

(iv) The (Nobel-prize winning) **Black-Scholes formula** for pricing a **European call option** which pays $\max(S_T - K, 0)$ dollars at time T in the future:

$$C(S,t) \quad = \quad S\Phi(d_1) - Ke^{-r\tau}\Phi(d_2)\,,$$

where $\tau = T - t$ and t is today,

$$d_1 \quad = \quad \frac{\log\frac{S}{K} + (r + \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}} \quad , \quad d_2 \quad = \quad d_1 - \sigma\sqrt{\tau}$$

and $\Phi(x) = \int_{-\infty}^{x} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz$ is the standard cumulative Normal distribution function.

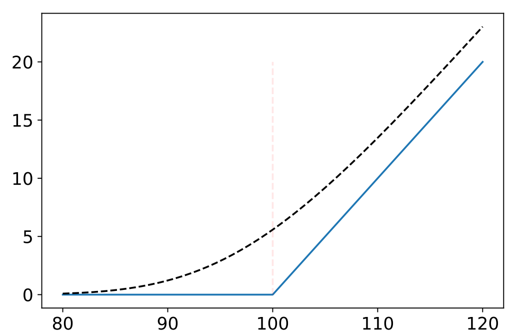

Figure 3: Solid blue line is the plot of the 'hockey stick' payoff of a European call option $\max(S - K, 0)$ for strike price $K = 100$ – payoff shown for $S \in (80, 120)$. Dotted black line is the Black-Scholes formula for $r = 0.03$, $\sigma = 0.1$, $t = 0$, and $T = 1$.

(v) We will show that the Black-Scholes call price $C(S,t)$ satisfies the Black-Scholes **partial differential equation**:

$$C_t + rSC_S + \frac{1}{2}\sigma^2 S^2 C_{SS} = rC$$

subject to the boundary condition $C(S,T) = \max(S-K,0)$ at the final maturity T, with solution $C = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}((S_T - K^+)$ where $dS_t = S_t(rdt + \sigma dW_t)$ under $\mathbb{Q}$. The important qualitative feature of the PDE and the solution is that it does not depend on the drift μ (similar to the binomial model in the 388 course). This is because we compute C using a hedging argument and not in absolute terms.

(vi) We will price derivatives using **Monte Carlo simulations** in Python. Many derivatives do not have a closed-form formula, thus, understanding how to price options using simulations is important. **Coding in Python will not be examined.**

(vii) Extra material (time-permitting) **not examinable**: Pricing and hedging **barrier options** under the Black-Scholes model. Barrier options either die or come to life if the share price hits a certain barrier level. These are more difficult and interesting to price and hedge since they are **path-dependent**.

Prerequisites for this course

- Some background in basic **probability**.

2 Probability essentials

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is composed of the following:

i) the state space: the set of all possible outcomes of the random experiment, usually denoted by Ω.

ii) the family of events: an event is a set of outcomes of the random experiment, which can be observed to hold or not to hold after the experiment. Mathematically, an event is a subset A of the state space Ω, that is $A \subseteq \Omega$. Often, the family of all events is denoted by $\mathcal{F}$.

iii) the probability: the probability is a number in $[0, 1]$ which is assigned to each event. The more likely an event is the closer is the assigned number to 1. Mathematically, the probability is a mapping $\mathbb{P} : \mathcal{F} \to [0, 1]$.

2.1 Kolmogorov's axioms

The set $\mathcal{A}$ of events is a collection of subsets of the state space Ω, that is $\mathcal{F} \subseteq \mathcal{P}(\Omega) := \{A \subseteq \Omega\}$.

Example 2.1 *If* $\Omega = \{1, 2, 3\}$ *then*

$$\mathcal{P}(\Omega) = \big\{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}, \emptyset\big\} \tag{2}$$

Definition 2.1 *Let* Ω *be a non-empty set and* $\mathcal{F}$ *be a family of subsets of* Ω*. Then* $\mathcal{F}$ *is called a* σ*-algebra of* Ω *if the following are satisfied:*

(i) $\Omega \in \mathcal{F}$

(ii) $A \in \mathcal{F}$ *then* $A^c \in \mathcal{F}$

(iii) if $A_1, A_2, \cdots \in \mathcal{F}$ *then* $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Example 2.2 *Some simple examples of a* σ*-algebra are the following:*

(i) $\{\emptyset, \Omega\}$ *– this is the trivial* σ*-algebra*

(ii) $\mathcal{P}(\Omega) = \{A \subseteq \Omega\}$ *– power set of* Ω

(iii) $\{\emptyset, \Omega, A, A^c\}$ *for* $A \subset \Omega$

In many cases it is not possible to describe explicitly all sets of a σ-algebra. However, for a given collection of sets one can define the smallest σ-algebra which contains the given collection of sets. For this purpose we need the following:

Theorem 2.1 *Let* $\{\mathcal{A}_i\}_{i \in I}$ *be a family of* σ*-algebras on* Ω *and* I *an arbitrary (not necessarily countable) index set. Then*

$$\bigcap_{i \in I} \mathcal{A}_i := \{A \subseteq \Omega : A \in \mathcal{A}_i \quad \forall i \in I\} \tag{3}$$

is a σ*-algebra on* Ω.

Theorem 2.1 enables us to define the concept of the generated σ-algebra: let $\mathcal{C}$ be a collection of subsets of Ω, that is $\mathcal{C} \subseteq \mathcal{P}(\Omega)$. The σ-algebra generated by $\mathcal{C}$ is denoted by $\sigma(\mathcal{C})$ and it is the unique σ-algebra satisfying:

(i) $\mathcal{C} \subseteq \sigma(\mathcal{C})$

(ii) if $\mathcal{A}$ is a σ-algebra of Ω with $\mathcal{C} \subseteq \mathcal{A}$ then $\sigma(\mathcal{C}) \subseteq \mathcal{A}$.

Example 2.3 *For a subset $A \subseteq \Omega$ we have $\sigma(\{A\}) = \{\emptyset, A, A^c, \Omega\}$.*

Definition 2.2 *Let $\Omega = \mathbb{R}$ or $\mathbb{R}^d$. The σ-algebra generated by all open sets in Ω is called the Borel σ-algebra in Ω and it is denoted by $\mathcal{B}(\mathbb{R})$ or $\mathcal{B}(\mathbb{R}^d)$.*

Definition 2.3 *A probability measure on a σ-algebra $\mathcal{F}$ is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ which satisfies:*

(i) $\mathbb{P}(\Omega) = 1$.

(ii) Each sequence $A_1, A_2, \dots \in \mathcal{F}$ of pairwise disjoint sets obeys

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)\,. \tag{4}$$

The sets $A_1, A_2, \dots \in \mathcal{F}$ are said to be pairwise disjoint if $A_i \cap A_j = \emptyset$ for $i \neq j$.
All ingredients for describing a random experiment are available.

Definition 2.4 *(Kolmogorov's axioms)*

(i) A measurable space $(\Omega, \mathcal{F})$ is a non-empty set Ω and a σ-algebra $\mathcal{F}$ of Ω.

(ii) A probability space is a triplet $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is a non-empty set, $\mathcal{F}$ is a σ-algebra of Ω and $\mathbb{P}$ is a probability measure on $\mathcal{F}$.

There are different interpretations of probabilities.[1] The most common one is the so-called frequentist probability, which interprets the probability $\mathbb{P}(A)$ of an event A as the limit of the relative frequency of the event A in a large number of trials. This corresponds to the examples of rolling a die and others. The mathematical approach, established by A. Kolmogorov in the 1930's, is purely axiomatic: within the mathematical framework we started here and will build up, one obtains many powerful results but all of them rely what the "user" actually models by the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. **In some cases, like rolling a die, the model $(\Omega, \mathcal{F}, \mathbb{P})$ is quite obvious, but in other cases, e.g. modelling a share price, it is much less. Even if one has introduced a model, which seems to present the real world very well, one should not forget that in the end it is just a model.**

Example 2.4 *Let us describe the probability space associated with the random experiment of throwing a dice. Here we have that*

(i) $\Omega = \{1, 2, 3, 4, 5, 6\}$

(ii) $\mathcal{F} = \mathcal{P}(\Omega)$

(iii) $\mathbb{P} : \mathcal{F} \to [0, 1]$ *is given by*

$$\mathbb{P}(A) = \frac{|A|}{6}\,, \qquad A \in \mathcal{F}\,. \tag{5}$$

[1] https://en.wikipedia.org/wiki/Probability_interpretations

2.2 Random variables, expectation, and other useful concepts

For completeness, we present the formal definition of a random variable.

Definition 2.5 *Let $(\Omega, \mathcal{F})$ and $(E, \mathcal{E})$ be two measurable spaces. A function $f : \Omega \to E$ is called $\mathcal{F} - \mathcal{E}$-measurable if*

$$f^{-1}(B) := \{\omega \in \Omega : f(\omega) \in B\} \in \mathcal{F} \quad \forall B \in \mathcal{E}\,. \tag{6}$$

A function $X : \Omega \to E$ is a random variable if it is $\mathcal{F} - \mathcal{E}$-measurable.

Thus, a random variable X is a quantity which depends on the outcome of the random experiment. Usually, we have $E = \mathbb{N}$ or $E = \mathbb{R}$. Whenever we use $\mathbb{R}$ in this course, the σ-algebra associated with it is $\mathcal{B}(\mathbb{R})$ and the measurable space is $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Condition (6) guarantees that for a random variable $X : \Omega \to E$ we can consider the probability $\mathbb{P}(\{X \in B\})$ of the event $\{X \in B\}$. We usually omit the brackets and only write $\mathbb{P}(X \in B)$. For these sets there are the following equivalent notations:

$$X^{-1}(B) = \{X \in B\} = \{\omega \in \Omega : X(\omega) \in B\}\,. \tag{7}$$

Remark 2.1 *In this course we will be working within a much simpler class of random variables, the so-called continuous random variables.*

Definition 2.6 *A continuous random variable $X : \Omega \to \mathbb{R}$ has the property that for any set A, the probability that X lies in the set A can be put in the form*

$$\mathbb{P}(X \in A) \quad = \quad \int_A f_X(u) du \tag{8}$$

where $f_X : \mathbb{R} \to [0, \infty)$ is the density of X.

Note that X is a random variable here and u is a dummy variable of integration. The case when $A = (-\infty, x]$ gives rise to the following definition.

Definition 2.7 *The distribution function of X is defined by*

$$F_X(x) \quad = \quad \mathbb{P}(X \in (-\infty, x]) \quad = \quad \mathbb{P}(X \leq x) \quad = \quad \int_{-\infty}^{x} f_X(u) du\,. \tag{9}$$

Then differentiating and using the fundamental theorem of calculus we see that

$$F'_X(x) \quad = \quad \frac{d}{dx} F_X(x) \quad = \quad f_X(x)\,.$$

Setting $x = \infty$ in (9) we see that

$$\mathbb{P}(X \leq \infty) \quad = \quad \mathbb{P}(X < \infty) \quad = \quad 1 \quad = \quad \int_{-\infty}^{\infty} f_X(x) dx\,,$$

i.e. the density of any continuous random variable has to integrate to 1. If we set A equal to the event $\{X \leq x\}$, then $A^c = \{X > x\}$, and $\mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c) = 1$, since A and A^c are two disjoint events such that $\mathbb{P}(A \cup A^c) = 1$ – see Definition 2.3.

The complementary cdf of X is defined as $\mathbb{P}(X > x)$, which is equal to $1-\mathbb{P}(A) = 1-\mathbb{P}(X \le x)$, and differentiating this expression with respect to x we see that

$$\frac{d}{dx}\mathbb{P}(X > x) \quad = \quad \frac{d}{dx}(1 - \mathbb{P}(X \le x)) \quad = \quad -f_X(x)$$

which will be used many times on the course. Note also that

$$\mathbb{P}(X = x) \quad = \quad \int_x^x f_X(u)du \quad = \quad 0$$

i.e. the probability that a continuous random variable X takes a particular value x is zero.

Example 2.5 *A standard uniform random variable U on $[0,1]$ has density $f_X(x) = 1$ for $x \in [0,1]$ and zero otherwise. The distribution function of U is obtained by integrating this density from 0 to x:*

$$\mathbb{P}(U \le x) \quad = \quad \int_0^x 1\, du \quad = \quad u|_{u=x} - u|_{u=0} \quad = \quad x \tag{10}$$

for $x \in [0,1]$. Note that the lower limit of integration is 0 not $-\infty$ here since the density of U is zero outside $[0,1]$. See first two panels in Figure 2.2 for a plot of the density and the cdf.

Example 2.6 *An exponential random variable X has density $f_X(x) = \lambda e^{-\lambda x}$ for $x \in (0,\infty)$ and zero otherwise, for some parameter $\lambda > 0$ (we say that X is an $\mathrm{Exp}(\lambda)$ random variable). The distribution function of X is again obtained by integrating this density from 0 to x as follows:*

$$\mathbb{P}(X \le x) \quad = \quad \int_0^x \lambda e^{-\lambda u} du \quad = \quad -e^{-\lambda u}|_{u=x} - (-e^{-\lambda u}|_{u=0}) \quad = \quad 1 - e^{-\lambda x}\,.$$

As a sanity check, we see that $\lim_{x\to 0}\mathbb{P}(X \le x) = \mathbb{P}(X \le 0) = 0$, and $\lim_{x\to+\infty}\mathbb{P}(X \le x) = \mathbb{P}(X < \infty) = 1$, as we would expect since $X \ge 0$ and $X < \infty$. See the two panels on the right of Figure 2.2 for a plot of the density and the cdf.

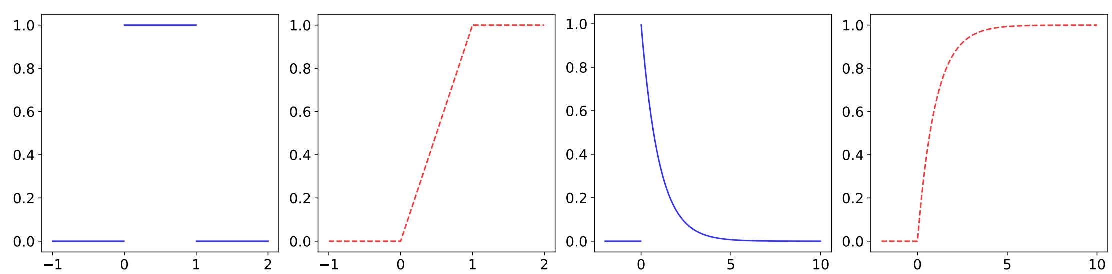

Figure 4: On the left we see the density of a standard $U[0,1]$ uniform random variable, and in the second from left panel we have plotted its distribution function $F_U(x) = x$. In the third panel we have plotted the density of an exponential random variable with parameter $\lambda = 1$, and on the right we have plotted its distribution function.

Definition 2.8 *The expectation of a continuous random variable X with density $f_X(x)$ is*

$$\mathbb{E}[X] \quad = \quad \int_{-\infty}^{\infty} x f_X(x) dx \,.$$

The expectation of $g(X)$ for some function g is given by

$$\mathbb{E}[g(X)] \quad = \quad \int_{-\infty}^{\infty} g(x) f_X(x) dx \tag{11}$$

Example 2.7 *Let X be a continuous random variable. For any constants $a, b \in \mathbb{R}$, setting $g(x) = ax + b$ in (11), we have*

$$\begin{aligned} \mathbb{E}[aX + b] &= \int_{-\infty}^{\infty} (ax + b) f_X(x) dx && (12) \\ &= a \int_{-\infty}^{\infty} x\, f_X(x) dx + b \int_{-\infty}^{\infty} f_X(x) dx && (13) \\ &= a\mathbb{E}[X] + b \,. && (14) \end{aligned}$$

Definition 2.9 *Let X be a random variable with expectation $\mathbb{E}(X) = \mu$. The variance of X is defined as*

$$\mathrm{Var}(X) \quad = \quad \mathbb{E}[(X - \mu)^2] \,, \tag{15}$$

which is equivalent to

$$\mathrm{Var}(X) = \mathbb{E}[X^2 - 2X\mu + \mu^2] = \mathbb{E}(X^2) - 2\mathbb{E}(X)\mu + \mu^2 = \mathbb{E}(X^2) \,-\, \mu^2 = \mathbb{E}(X^2) \,-\, \mathbb{E}^2(X) \,. \tag{16}$$

Example 2.8 *From the definition of variance we see that for any constants $a, b \in \mathbb{R}$*

$$\mathrm{Var}(aX) \quad = \quad \mathbb{E}[(aX)^2] \,-\, (\mathbb{E}[aX])^2 \quad = \quad a^2\mathbb{E}[X^2] \,-\, a^2(\mathbb{E}[X])^2 \quad = \quad a^2\mathrm{Var}(X) \,,$$

where we are using (14) to obtain the middle equality. Similarly,

$$\begin{aligned} \mathrm{Var}(X + b) &= \mathbb{E}[(X + b)^2] \,-\, (\mathbb{E}[X + b])^2 = \mathbb{E}[X^2 + 2Xb + b^2] \,-\, \big((\mathbb{E}[X])^2 + 2b\mathbb{E}[X] + b^2\big) && (17) \\ &= \mathbb{E}[X^2] + 2\,b\,\mathbb{E}[X] + b^2 - (\mathbb{E}[X])^2 - 2\,b\,\mathbb{E}[X] - b^2 && (18) \\ &= \mathrm{Var}(X) \,. \end{aligned}$$

Definition 2.10 *Let X, Y be two random variables in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that X and Y are independent if*

$$\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x \cap Y \leq y) = \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y) \,. \tag{19}$$

Definition 2.11 *Let X, Y be two continuous random variables. The joint cumulative distribution function is defined as*

$$F_{X,Y}(x, y) := \mathbb{P}(X \leq x, Y \leq y) \,. \tag{20}$$

Proposition 2.2 *The joint cumulative distribution function $F_{X,Y}(x,y)$ of X, Y satisfy*

(i) $F_X(x) = F_{X,Y}(x, \infty)$

(ii) $F_Y(y) = F_{X,Y}(\infty, y)$

(iii) $f_{X,Y}(x,y) = \partial_x \, \partial_y F_{X,Y}(x,y)$

Example 2.9 *Let X and Y be two independent continuous random variables with densities $f_X(x)$ and $f_Y(y)$ respectively, then $f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$ and $\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)]$.*

Proof. The fact that $f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$ follows from

$$F_{X,Y}(x,y) = F_X(x)\, F_Y(y)\,, \tag{21}$$

by definition and taking partials w.r.t. x and y. Furthermore,

$$\begin{aligned}
\mathbb{E}[g(X)h(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)\, h(y)\, f_{X,Y}(x,y)\, dxdy &(22)\\
&= \int_{-\infty}^{\infty} g(x) f_X(x) dx \int_{-\infty}^{\infty} h(y) f_Y(y) dy &(23)\\
&= \mathbb{E}[g(X)]\, \mathbb{E}[h(Y)]\,. &(24)
\end{aligned}$$

∎

Definition 2.12 *Let X and Y be random variables. We define the covariance between X and Y by*

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)]\,, \tag{25}$$

where $\mu_X := \mathbb{E}(X)$ and $\mu_Y := \mathbb{E}(Y)$.

Example 2.10 *For two random variables X and Y we have that*

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) \,+\, \mathrm{Var}(Y) \,+\, 2\mathrm{Cov}(X,Y)\,.$$

Furthermore, if X and Y are independent then $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Proof. By definition

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathbb{E}((X+Y)^2) - (\mathbb{E}(X+Y))^2\\
&= \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) \,-\, (\mathbb{E}(X)^2 + \mathbb{E}(Y)^2 + 2\mathbb{E}(X)\mathbb{E}(Y))\\
&= \mathrm{Var}(X) \,+\, \mathrm{Var}(Y) \,+\, 2\mathrm{Cov}(X,Y)\,.
\end{aligned}$$

If X and Y are independent random variables, then $\mathbb{E}((X-\mu_X)(Y-\mu_Y)) = \mathbb{E}(X-\mu_X)\mathbb{E}(Y-\mu_Y) = 0 \times 0 = 0$, so $\mathrm{Var}(X+Y)$ simplifies to

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) \,+\, \mathrm{Var}(Y)\,.$$

∎

Definition 2.13 *We say Z has a standard $N(0,1)$ Normal distribution, if the density function $f_Z(x)$ is given by*

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \tag{26}$$

for $x \in \mathbb{R}$.

Remark 2.2 *Let $Z \sim N(0,1)$. We have that $\mathbb{E}[Z] = 0$ and $\mathrm{Var}(Z) = 1$. The density of Z is a "bell-shaped" function – see Figure 5, and the distribution function of Z is the "s-shaped" function:*

$$\mathbb{P}(Z \leq x) \quad = \quad \Phi(x) \quad := \quad \int_{-\infty}^{x} f_Z(z)dz \tag{27}$$

and as before

$$\Phi'(x) \quad = \quad f_Z(x)\,.$$

Remark 2.3 *The function $f_Z(x)$ above will be used often throughout the course, we will have a special notation for it:*

$$n(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,. \tag{28}$$

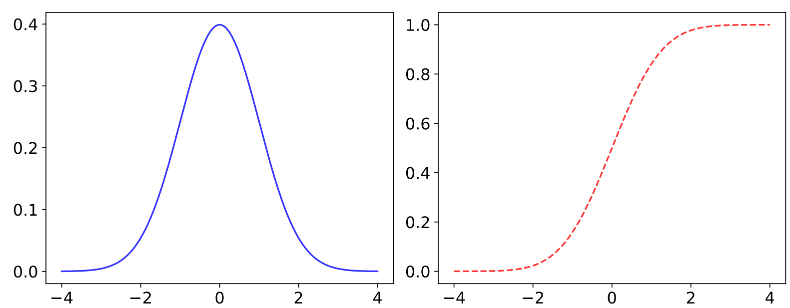

Figure 5: Here we have plotted the density $f_Z(x)$ (left) and the distribution function $\Phi(x)$ (right) for the standard Normal distribution.

Proving that $\int_{-\infty}^{\infty} n(z)dz = 1$ is not trivial, and requires working in polar coordinates with two i.i.d. (independent and identically distributed) standard Normal random variables. The quantity $\Phi(x)$ cannot be computed exactly, but there are useful asymptotic formula for $\Phi(x)$ when x is small or x is large, or we can look up Φ in tables. The function $\Phi^c(x)$ is defined as $\Phi^c(x) = 1 - \Phi(x)$.

Normal distributions are ubiquitous. The next theorem explains why.

Theorem 2.3 *Let X_1, X_2, X_3, $\ldots$ be i.i.d. samples from a distribution with mean μ and variance $\sigma^2 < \infty$. For $n \in \mathbb{N}$ let*

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i\,, \tag{29}$$

then,

$$Z = \lim_{n\to\infty} \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \tag{30}$$

has a standard Normal distribution $N(0,1)$.

Definition 2.14 *A general Normal random variable* $X \sim N(\mu, \sigma^2)$ *with mean* μ *and variance* σ^2 *has density given by*

$$f_X(x) = \frac{1}{\sqrt{2\,\pi}\,\sigma}\, e^{-\frac{(x-\mu)^2}{2\,\sigma^2}}\,, \tag{31}$$

for $x \in \mathbb{R}$.

Definition 2.15 *The moment generating function (mgf) of a random variable* X *is defined as*

$$M_X(p) := \mathbb{E}(e^{pX}) \quad = \quad \int_{-\infty}^{\infty} e^{px} f_X(x) dx$$

for $p \in \mathbb{R}$, *where we are using* (11) *for the final equality.*

Proposition 2.4 *If* $X \sim N(\mu, \sigma^2)$, *then the mgf of* X *is given by*

$$M_X(p) = e^{\,\mu p + \frac{1}{2}\sigma^2 p^2}\,. \tag{32}$$

Proof. We have that

$$\begin{aligned}
\mathbb{E}(e^{pX}) &= \int_{-\infty}^{\infty} e^{p\,x} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\,\sigma^2}} dx &(33)\\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{x^2-2\,x(\mu+p\sigma^2)+(\mu+p\sigma^2)^2-(\mu+p\sigma^2)^2+\mu^2}{2\,\sigma^2}} dx &(34)\\
&= e^{-\frac{-(\mu+p\sigma^2)^2+\mu^2}{2\,\sigma^2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{x^2-2\,x(\mu+p\sigma^2)+(\mu+p\sigma^2)^2}{2\,\sigma^2}} dx &(35)\\
&= e^{-\frac{-2\,p\mu\sigma^2-p^2\sigma^4}{2\,\sigma^2}} &(36)\\
&= e^{\,\mu p + \frac{1}{2}\sigma^2 p^2}\,. &(37)
\end{aligned}$$

■

Moment generating functions are also used to identify the distribution associated with a given random variable.[2] This is because the mgf uniquely determines the distribution. Below, we give two examples of how this can be used. Both are very important for this course.

Proposition 2.5 *Let* $X \sim N(\mu, \sigma^2)$, *and let*

$$Z = \frac{X-\mu}{\sigma}\,, \tag{38}$$

then

$$M_Z(p) = e^{\frac{1}{2}p^2}\,. \tag{39}$$

[2]It is important not to confuse random variables with distributions. For example, if $X \sim N(0,1)$, both X and $-X$ have the same distribution (i.e., $N(0,1)$ distribution) but the outcomes of the random variables are different (one is minus the other).

Proof. The moment generating function of Z is given by

$$M_Z(p) = \mathbb{E}\left[e^{p\,Z}\right] = \mathbb{E}\left[e^{p\,\frac{X-\mu}{\sigma}}\right] \tag{40}$$

$$= e^{-p\,\frac{\mu}{\sigma}}\,\mathbb{E}\left[e^{\frac{p}{\sigma}\,X}\right] \tag{41}$$

$$= e^{-p\,\frac{\mu}{\sigma}}\,M_X\left(\frac{p}{\sigma}\right). \tag{42}$$

Note how we went from M_Z to M_X which in turn gives

$$M_Z(p) = e^{-p\,\frac{\mu}{\sigma}}\,e^{\mu\,\frac{p}{\sigma}+\frac{1}{2}\,\sigma^2\,\frac{p^2}{\sigma^2}} \tag{43}$$

$$= e^{\frac{1}{2}p^2}. \tag{44}$$

■

Given that moment generating functions characterise distributions, the calculations in Proposition 2.5 are a proof of the following useful result.

Corollary 2.6 *Let $X \sim N(\mu, \sigma^2)$ and define*

$$Z = \frac{X-\mu}{\sigma}, \tag{45}$$

then, $Z \sim N(0,1)$.

Proposition 2.7 *Let $X \sim N(\mu, \sigma^2)$, then*

$$\mathbb{P}(X \le x) = \Phi\left(\frac{x-\mu}{\sigma}\right). \tag{46}$$

Proof. Take $Z = (X-\mu)/\sigma$, then

$$\mathbb{P}(X \le x) \quad = \quad \mathbb{P}\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) \quad = \quad \mathbb{P}\left(Z \le \frac{x-\mu}{\sigma}\right) \quad = \quad \Phi(z)$$

where $z = \frac{x-\mu}{\sigma}$ and $\Phi(x)$ is the cumulative distribution function of the standard normal random variable – see (27). ■

Proposition 2.8 *Let X and Y be independent random variables with $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$. Define $S = X + Y$, then $S \sim N(\mu_1 + \mu_2,\, \sigma_1^2 + \sigma_2^2)$.*

Proof. Note that

$$\mathbb{E}\left[e^{p\,S}\right] = \mathbb{E}\left[e^{p\,X}\,e^{p\,Y}\right] = \mathbb{E}\left[e^{p\,X}\right]\mathbb{E}\left[e^{p\,Y}\right] \tag{47}$$

$$= e^{\mu_1\,p+\frac{1}{2}\sigma_1^2p^2}\,e^{\mu_2\,p+\frac{1}{2}\sigma_2^2p^2} \tag{48}$$

$$= e^{(\mu_1+\mu_2)\,p+\frac{1}{2}(\sigma_1^2+\sigma_2^2)\,p^2}, \tag{49}$$

and from Proposition 2.4 we know this is a Normal random variable with mean $\mu_1+\mu_2$ and variance $\sigma_1^2+\sigma_2^2$. Thus,

$$S \sim N(\mu_1+\mu_2,\, \sigma_1^2+\sigma_2^2). \tag{50}$$

■

Proposition 2.9 *If $X \geq 0$ and $\mathbb{E}\left[e^{pX}\right] < \infty$ for p in some open interval $I = (p_*, p^*)$ which contains zero, then*

$$M^{(n)}(0) = \mathbb{E}[X^n]\,. \tag{51}$$

Proof. From a result called Fubini's theorem, we can interchange the expectation and the integral if the integrand is non-negative, so

$$\int_0^p \mathbb{E}\left[Xe^{qX}\right] dq \quad = \quad \mathbb{E}\left[\int_0^p Xe^{qX} dq\right] \quad = \quad \mathbb{E}\left[e^{qX}|_{q=0}^{q=p}\right] \quad = \quad \mathbb{E}\left[e^{pX} - 1\right]\,.$$

Hence by inspection of the left hand side and the fundamental theorem of calculus, we see that $\frac{d}{dp}\mathbb{E}[e^{pX} - 1] = \frac{d}{dp}\mathbb{E}[e^{pX}] = \mathbb{E}[Xe^{pX}]$. Setting $p = 0$ we see that

$$\frac{d}{dp}\mathbb{E}[e^{pX}]|_{p=0} = \mathbb{E}(X)\,.$$

By repeating this procedure, we can show that

$$M^{(n)}(p) := \frac{d^n}{dp^n}\mathbb{E}\left[e^{pX}\right] \quad = \quad \mathbb{E}\left[X^n e^{pX}\right]$$

so $M^{(n)}(0) = \mathbb{E}[X^n]$, which is the nth moment of X. This is why M is called the moment generating function. ■

Note: To extend to a real-valued random variable X which may be negative, we have to check that $\int_0^p \mathbb{E}(|Xe^{qX}|)dq < \infty$ to justify use of Fubini.

Proposition 2.10 *If $X_1, ..., X_n$ is a sequence of i.i.d. random variables, then*

$$\mathbb{E}\left[e^{p(X_1+...+X_n)}\right] \quad = \quad \mathbb{E}\left[e^{pX_1}\right]\cdots\mathbb{E}\left[e^{pX_n}\right] \quad = \quad \mathbb{E}\left[e^{pX_1}\right]^n\,.$$

2.3 Stochastic processes

Definition 2.16 *Let I be a subset of $[0, \infty)$. A stochastic process with values in $\mathbb{R}$ is a family $(X_t : t \in I)$ of random variables $X_t : \Omega \to \mathbb{R}$ for each $t \in I$.*

In this course we consider stochastic processes in continuous time, that is $I = [0, T]$ for a constant $T > 0$. Sometimes we use the notation $(X_t)_{t\in I}$ instead of $(X_t : t \in I)$; both mean the same.

There are at least two different perspectives on stochastic processes:

- for each fixed $t \in I$ the object $X_t : \Omega \to \mathbb{R}$ is a random variable and the stochastic process $(X_t : t \in I)$ might be considered as an ordered family of random variables;
- for each fixed $\omega \in \Omega$ the collection $\{X_t(\omega) : t > 0\}$ is a function

$$t \to X_t(\omega)\,. \tag{52}$$

This mapping is called a path or a trajectory of X.

If $\mathbb{P}$-almost all (for short: $\mathbb{P}$-a.a.) paths of a stochastic process have a certain property, then we describe the stochastic process by this property, e.g. a continuous stochastic process $(X(t) : t > 0)$ means that for $\mathbb{P}$-a.a. $\omega \in \Omega$ its trajectories $t \to X_t(\omega)$ are continuous. Here for $\mathbb{P}$-a.a. $\omega \in \Omega$ means that there exists a set $\Omega_0 \subseteq \Omega \in \mathcal{F}$ with $\mathbb{P}(\Omega_0) = 1$ such that the property holds for all $\omega \in \Omega_0$, e.g. the trajectories $t \to X_t(\omega)$ are continuous for all $\omega \in \Omega_0$.

Example 2.11 *Let $X_1, X_2, \dots$ be independent, identically distributed random variables with $\mathbb{P}(X_i = 1) = p$ and $\mathbb{P}(X_i = -1) = 1 - p$, for some fixed value $p \in (0, 1)$, $i \in \mathbb{N}$. Define $R_t = 0$ for $t \in [0, 1)$ and for each $t \geq 1$*

$$R_t := \sum_{i=1}^{\lfloor t \rfloor} X_i \,, \tag{53}$$

where $\lfloor t \rfloor$ denotes the largest integer smaller than t. It follows that $(R_t : t > 0)$ is a stochastic process, the so-called random walk.

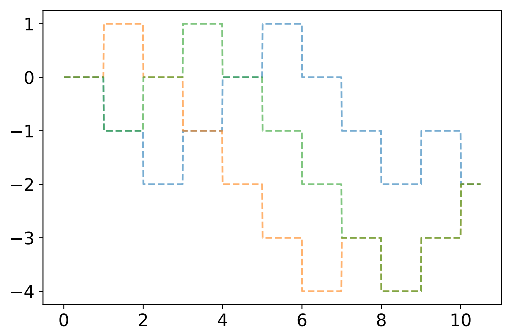

Figure 6: Three simulations of a random walk with $p = 0.4$.

Definition 2.17 *Let $(\Omega, \mathcal{F})$ be a measurable space. A family $\{\mathcal{F}_t\}_{t \in I}$ of σ-algebras $\mathcal{F}_t \subseteq \mathcal{F}$ for all $t \in I$ with*

$$\mathcal{F}_s \subseteq \mathcal{F}_t \quad \text{for all } 0 \leq s \leq t \tag{54}$$

is called a filtration.

Definition 2.18 *A stochastic process $(X_t : t \in I)$ is called adapted with respect to a filtration $\{\mathcal{F}_t\}_{t \in I}$ if X_t is $\mathcal{F}_t$-measurable for every $t \in I$.*

You might think of a filtration as the description of the information available at different times. The σ-algebra $\mathcal{F}_s$ of a filtration $\{\mathcal{F}_t\}_{t \in I}$ represents the information which is available at time s. One can think that the random outcome $\omega \in \Omega$ is already specified but we are only told at time s for all sets in the σ-algebra $\mathcal{F}_s$ whether this ω is in the set or not. The more sets there are in $\mathcal{F}_s$, the more information we obtain of an $\mathcal{F}_s$-measurable random variable. If $(X_t : t > 0)$ is an adapted stochastic process this means that the random function $t \to X_t(\omega)$ is already specified on the interval $[0, \infty)$ (by fixing $\omega \in \Omega$) but we know at time s only the values of the function on the interval $[0, s]$ but not on (s, ∞).

Most often one assumes the filtration which is generated by the process X itself, that is at a time t the σ-algebra $\mathcal{F}_t$ contains all information which is decoded by X restricted to the interval $[0,t]$. This is described by

$$\mathcal{F}_t^X := \sigma\left(X_s \,:\, s \in [0,t]\right) \tag{55}$$

which is

$$\mathcal{F}_t^X := \sigma\left((X_s)^{-1}([a,b]) \,:\, s \in [0,t],\, -\infty < a \leq b < \infty\right) , \tag{56}$$

and it means "the smallest σ-algebra" containing all the pre-images

$$(X_s)^{-1}([a,b]) := \{\omega \in \Omega \,:\, X_s(\omega) \in [a,b]\} , \tag{57}$$

for all $s \in [0,t]$ and all $a, b \in \mathbb{R}$.

There are many more advanced concepts regarding filtrations that we omit in this course. For the curious mind, please see Chapters 1-2 in Shreve (2004).

2.4 Conditioning

Definition 2.19 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with $A, B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$, then, the probability of A given B, denoted by $\mathbb{P}(A|B)$ is defined as*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} . \tag{58}$$

In the context of continuous random variables, if X and Y are two continuous random variables with densities $f_X(x)$ and $f_Y(y)$, then, the conditional density of X given Y

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \qquad \text{if } f_Y(y) > 0 , \tag{59}$$

and zero otherwise.

In this case, the conditional expectation of $g(X)$ given $Y = y$ is given by

$$\mathbb{E}[g(X)|Y = y] = \int_{-\infty}^{\infty} g(x)\, f_{X|Y=y}(x)\, dx , \tag{60}$$

for a measurable function g.

Lastly, if X is independent of Y, that is, if $f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$, then one can check that

$$\mathbb{E}[g(X)|Y = y] = \mathbb{E}[g(X)] , \tag{61}$$

i.e., one can drop Y from the conditioning.

Conditional expectations have another property that sometimes is referred to as "taking out what is known"; in the context of continuous random variables this can be written as

$$\mathbb{E}[f(X)\, g(Y) \,|\, Y] = g(Y)\, \mathbb{E}[f(X)|Y] , \tag{62}$$

where we think of Y as being known, therefore $g(Y)$ is also known and can be taken out of the expectation. Rigorously, for the above to hold one needs the various variables to be integrable (which we assume) – see §9.7 in Williams (1991) for details.

Definition 2.20 *Let $(X_t)_{t\in I}$ be a stochastic process adapted to the filtration $(\mathcal{F}_t)_{t\in I}$. We say that X has the Markov property if for any $0 \le s \le T$ and measurable function g we have that*

$$\mathbb{E}[g(X_T) \,|\, \mathcal{F}_s] = \mathbb{E}[g(X_T) \,|\, X_s]\,. \tag{63}$$

Throughout this course all stochastic processes will be assumed to have the Markov property. Equation (63) can be thought as follows: from all the information of the trajectory of X up to time s, it suffices to take X_s to have all the possible information about the future.

2.5 Other interesting concepts

Proposition 2.11 ***Strong law of large numbers*** *(SLLN). Let $X_1, X_2, \ldots$ denote an infinite sequence of independent identically distributed (i.i.d) random variables with $\mathbb{E}(X_i) = \mu$ and $\mathbb{E}(X_i^4) < \infty$ for all $i = 1, 2, \ldots$. Let $S_n = \sum_{i=1}^n X_i$. Then the SLLN says that*

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{S_n}{n} = \mu\right) \quad = \quad 1$$

i.e. the sample average $\frac{S_n}{n}$ tends to the true expectation μ of X_i as $n \to \infty$, as we might expect. This result is the cornerstone of Monte Carlo simulation, namely that if μ is unknown, we can estimate μ with $\hat{\mu}_n = \frac{S_n}{n}$ for n large – see Chapter 7 of Williams (1991) for a proof.

Proposition 2.12 ***Transformations of a 1-d random variable****. Let X be a continuous random variable with density $f_X(x)$, and let $Y = g(X)$, where g is differentiable and strictly increasing, which implies that $g'(x) > 0$ for all $x \in \mathbb{R}$, which further implies that g has a unique inverse g^{-1} such that $g^{-1}(g(x)) = x$. Then the distribution function of Y is*

$$\mathbb{P}(Y \le y) = F_X(h(y))F_X(h(y)) \tag{64}$$

where $h(y) = g^{-1}(y)$. The density function of Y is

$$f_Y(y) = h'(y) f_X(h(y))\,. \tag{65}$$

Proof. We have that

$$\mathbb{P}(Y \le y) \quad = \quad \mathbb{P}(X \le g^{-1}(y)) \quad = \quad F_X(g^{-1}(y)) \quad = \quad F_X(h(y))\,,$$

where $h(y) = g^{-1}(y)$. Differentiating with respect to y, we obtain the density of Y:

$$f_Y(y) \quad = \quad \frac{d}{dy}\mathbb{P}(Y \le y) \quad = \quad h'(y)F_X'(h(y)) \quad = \quad h'(y) f_X(h(y))$$

using the chain rule. ■

Proposition 2.13 ***Generating correlated Normal random variables****. Let Z_1, Z_2 be two i.i.d $N(0,1)$ random variables and*

$$\begin{aligned} X &= Z_1 \\ Y &= \rho Z_1 + \sqrt{1-\rho^2} Z_2 \end{aligned} \tag{66}$$

for some ρ with $-1 \le \rho \le 1$. Then, both X and Y are standard $N(0,1)$ and

$$\mathrm{Corr}(X, Y) = \rho\,. \tag{67}$$

Proof. Trivially X is $N(0,1)$. It follows from Proposition 2.8 that Y is a $N(0,1)$. A precise calculation of the variance of the sum is

$$\begin{aligned}\mathrm{Var}(Y) &= \mathrm{Var}(\rho Z_1 + \sqrt{1-\rho^2} Z_2) = \mathrm{Var}(\rho Z_1) + \mathrm{Var}(\sqrt{1-\rho^2} Z_2) \\ &= \rho^2 \mathrm{Var}(Z_1) + (1-\rho^2)\mathrm{Var}(Z_2) \\ &= \rho^2 + 1 - \rho^2 = 1\,.\end{aligned} \tag{68}$$

Then,

$$\mathrm{Corr}(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\mathbb{E}((X-\mu_X)(Y-\mu_Y))}{\sigma_X \sigma_Y} = \mathbb{E}(XY) = \rho \tag{69}$$

where $\mu_X = \mathbb{E}(X) = 0$, $\mu_Y = \mathbb{E}(Y) = 0$, and σ_X, σ_Y denote the standard deviations of X and Y (which in this case are also both 1 since Z_1 and Z_2 are standard Normal RVs).

To verify the last equality we see that

$$\mathbb{E}(XY) = \mathbb{E}(Z_1(\rho Z_1 + \sqrt{1-\rho^2} Z_2)) = \rho \mathbb{E}(Z_1^2) + \sqrt{1-\rho^2}\, \mathbb{E}(Z_1)\mathbb{E}(Z_2) = \rho \mathbb{E}(Z_1^2) = \rho$$

since $\mathrm{Var}(Z_1) = \mathbb{E}(Z_1^2) - \mathbb{E}(Z_1)^2 = \mathbb{E}(Z_1^2)$ as $\mathbb{E}(Z_1) = 0$, which verifies (69). To show that Y is a normal random variable we recall Proposition 2.8. ■

2.6 Example questions

The questions below are examples of what I would expect you to be able to solve in the context of an examination.

1. For a continuous random variable X with density $f_X(x)$, prove that $\mathbb{E}(1_A(X)) = \mathbb{P}(X \in A)$ for any set A, where $1_A(X) = 1$ if $X \in A$ and zero otherwise. $1_A(X)$ is known as the indicator function, and will be used many times in the course, and we sometimes abbreviate this to just 1_A.

Solution.

$$\mathbb{E}[1_A(X)] = \int_{-\infty}^{\infty} 1_A(x) f_X(x) dx = \int_A f_X(x) dx = \mathbb{P}(X \in A)\,.$$

2. Let W_t be a random function such that $W_t \sim N(0,t)$; Brownian motion has this nice property as we shall see in Section 3. Compute $\mathbb{P}(W_t > x)$.

Solution.

$$\mathbb{P}(W_t > x) = \mathbb{P}\left(\frac{W_t - 0}{\sqrt{t}} > \frac{x-0}{\sqrt{t}}\right) = \mathbb{P}\left(Z > \frac{x}{\sqrt{t}}\right) = \Phi^c\left(\frac{x}{\sqrt{t}}\right)$$

since $Z = (W_t - 0)/\sqrt{t}$ is a standard $N(0,1)$ random variable. The general rule here is: do to one side what you do to the other side.

3. Simulating random variables with a given distribution. Let X be a random variable with a continuous strictly increasing distribution function $F_X(x)$. What is the distribution of $F_X^{-1}(U)$, where U is a standard Uniform random variable on $[0,1]$?

Solution.

$$\mathbb{P}(F_X^{-1}(U) \leq x) \quad = \quad \mathbb{P}(U \leq F_X(x)) \quad = \quad F_X(x) \tag{70}$$

since $F_X(F_X^{-1}(x)) = x$, i.e. $F_X^{-1}(U) \sim X$, where we have used that $\mathbb{P}(U \leq x) = x$ from (10), i.e. $F_X^{-1}(U)$ has the same distribution as X. This is how we typically generate a random variable with a given distribution in practice on a computer.

Similarly, we can compute the distribution function of $F_X(X)$ as

$$\mathbb{P}(F_X(X) \leq x) \quad = \quad \mathbb{P}(X \leq F_X^{-1}(x)) \quad = \quad F_X(F_X^{-1}(x)) \quad = \quad x \tag{71}$$

so we see that $F_X(X)$ has the same distribution function as a $U[0,1]$ random variable – see Equation (10) above, so $F_X(X) \sim U[0,1]$.

3 Brownian motion

Definition 3.1 *A continuous-time stochastic process $(W_t)_{t\geq 0}$ is said to be a standard one-dimensional Brownian motion if it satisfies the following four conditions:*

(i) $W_0 = 0$.

(ii) W *has independent increments, i.e.*

$$W_{t_2} - W_{t_1},\ W_{t_3} - W_{t_2},\ \dots,\ W_{t_n} - W_{t_{n-1}}$$

are independent for all $0 \leq t_1 < t_2 < \dots < t_n$.

(iii) The increments are normally distributed: $W_t - W_s \sim N(0, t-s)$ *for all* $0 \leq s < t$.

(iv) W_t *is continuous as a function of* t *almost surely (i.e. with probability one). This means that there is* $\Omega_0 \subseteq \Omega$ *with* $\mathbb{P}(\Omega_0) = 1$ *such that for* $\omega \in \Omega_0$ *the map* $t \to W_t(\omega)$ *is continuous.*

Remark 3.1 *It is not immediately obvious that we can rigorously construct a process W which satisfies these four properties. See Cohen and Elliott (2015) pages 120–124 for Lévy's construction of Brownian motion. Here, we will study the Euler method (see also numerical simulation on next page to see what Brownian motion looks like).*

Remark 3.2 *We can work with $\mathcal{F}_t = \sigma(W_s : 0 \leq s \leq t)$, or often it is more convenient to work with a larger filtration called the augmented filtration – see Shreve (2004) for more information; do not worry about the details of this filtration for this course.*

3.1 Elementary properties of Brownian motion

For this course, the third property is the most important to remember.

Proposition 3.1 *Let $(W_t)_{t\geq 0}$ be a Brownian motion and let $u > 0$, then*

$$W_u \quad \sim \quad N(0, u) \tag{72}$$

and therefore

$$\mathbb{E}[W_u] = 0\,, \tag{73}$$

$$\mathrm{Var}(W_u) = u\,. \tag{74}$$

Proof. The result in (72) is a consequence of the third property for $t = u$ and $s = 0$ together with the property that $W_0 = 0$. ■

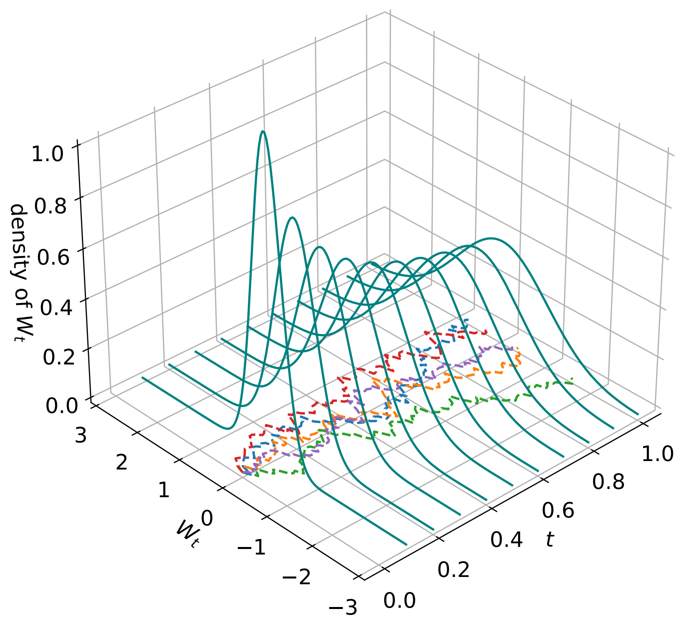

Figure 7: Five paths of Brownian motion and its density at various points in time.

Proposition 3.2 *Let $(W_t)_{t\geq 0}$ be a Brownian motion. Given that $W_t \sim N(0,t)$ we have that*

$$\mathbb{P}(W_t > x) = \Phi^c\left(\frac{x}{\sqrt{t}}\right) \qquad \textit{and} \qquad \mathbb{P}(W_t \leq x) = \Phi\left(\frac{x}{\sqrt{t}}\right). \tag{75}$$

Proof. We have that

$$\mathbb{P}(W_t > x) = \mathbb{P}\left(\frac{W_t - 0}{\sqrt{t}} > \frac{x-0}{\sqrt{t}}\right) = \mathbb{P}\left(Z > \frac{x}{\sqrt{t}}\right) = \Phi^c\left(\frac{x}{\sqrt{t}}\right) \tag{76}$$

where $\Phi^c(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$, and we are using that $Z = (W_t - 0)/\sqrt{t}$ is a standard Normal random variable. The second equality follows from the identity $\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$. We will use this equation many times. ■

Proposition 3.3 *Let $(W_t)_{t\geq 0}$ be a Brownian motion, then*

$$\mathbb{E}(W_s W_t) = \min(s,t). \tag{77}$$

Proof. Let $0 \leq s \leq t$. Then

$$\mathbb{E}(W_s W_t) \quad = \quad \mathbb{E}(W_s(W_s + W_t - W_s)) \quad = \quad \mathbb{E}(W_s^2) \quad = \quad \mathrm{Var}(W_s) \quad = \quad s$$

since $\mathbb{E}(W_s(W_t - W_s)) = \mathbb{E}((W_s - W_0)(W_t - W_s)) = \mathbb{E}(W_s - W_0)\mathbb{E}(W_t - W_s) = 0$. This means that in general, for $s, t \geq 0$

$$R(s,t) \quad := \quad \mathbb{E}(W_s W_t) \quad = \quad \min(s,t)\,. \tag{78}$$

This is known as the **covariance function** of Brownian motion. ■

Proposition 3.4 *Let $(W_t)_{t\geq 0}$ be a Brownian motion, and let $0 < s < t$, then*

$$\mathbb{P}\left(W_t \leq x | W_s = y\right) = \Phi\left(\frac{x-y}{\sqrt{t-s}}\right)\,. \tag{79}$$

Proof. We have that

$$\begin{aligned}
\mathbb{P}\left(W_t \leq x | W_s = y\right) &= \mathbb{P}\left(W_t - W_s + W_s \leq x | W_s = y\right) && (80)\\
&= \mathbb{P}\left(W_t - W_s + y \leq x | W_s = y\right) && (81)\\
&= \mathbb{P}\left(W_t - W_s \leq x - y | W_s = y\right) && (82)\\
&= \mathbb{P}\left(W_t - W_s \leq x - y\right)\,, && (83)
\end{aligned}$$

where in the last equality follows as a consequence of $W_t - W_s$ being independent from W_s (Definition 3.1 (*ii*)). Lastly,

$$\begin{aligned}
\mathbb{P}\left(W_t \leq x | W_s = y\right) &= \mathbb{P}\left(\frac{W_t - W_s}{\sqrt{t-s}} \leq \frac{x-y}{\sqrt{t-s}}\right) && (84)\\
&= \Phi\left(\frac{x-y}{\sqrt{t-s}}\right)\,. && (85)
\end{aligned}$$

■

Corollary 3.5 *Let $(W_t)_{t\geq 0}$ be a Brownian motion, and let $0 < s < t$, then*

$$\mathbb{E}\left[W_t \,|\, W_s = y\right] = y\,. \tag{86}$$

Corollary 3.6 *Let $(W_t)_{t\geq 0}$ be a Brownian motion, and let $0 < s < t$, then*

$$f_{W_t \,|\, W_s = y}(x) = \frac{1}{\sqrt{2\,\pi(t-s)}} e^{-\frac{(x-y)^2}{2\,(t-s)}}\,. \tag{87}$$

Proposition 3.7 *Let $(W_t)_{t\geq 0}$ be a standard Brownian motion and let $\lambda \neq 0$, then the process $(Z_t)_{t\geq 0}$ defined by*

$$Z_t = \lambda\, W_{t/\lambda^2}\,, \tag{88}$$

is a standard Brownian motion.

Proof. Homework 3. ■

Corollary 3.8 *Let $(W_t)_{t\geq 0}$ be a standard Brownian motion then $(-W_t)_{t\geq 0}$ is also a standard Brownian motion.*

3.2 Lack of differentiability

It turns out that Brownian motion is nowhere differentiable. More precisely, we have a relevant result; see Karatzas and Shreve (1991) pages 109–111 for a proof.

Theorem 3.9 *For almost every $\omega \in \Omega$, the Brownian sample path $t \to W_t(\omega)$ is nowhere differentiable.*

Remark 3.3 *Brownian motion is not differentiable, but from the fourth property in the definition we know that W is continuous a.s. (so it does not have jumps), and (using something called the Kolmogorov continuity theorem) we can make a stronger statement that W is α-Hölder continuous for $\alpha \in (0, \frac{1}{2})$, i.e. for $0 \le s \le t \le T$*

$$|W_t - W_s| \quad \le \quad c_1 |t - s|^{\alpha}$$

for some random constant c_1 which depends on $(W_t)_{0 \le t \le T}$ which is finite almost surely (a.s.). Note this implies that W is continuous a.s. since as $t - s \to 0$, $|W_t - W_s| \to 0$ because $\alpha > 0$.

3.3 Constructing and simulating Brownian motion

Let Z_i be a sequence of standard i.i.d. $N(0,1)$ random variables. Then we can approximate Brownian motion numerically as follows: fix a small step size $\Delta t > 0$, set $W_0^n = 0$, and then iteratively define

$$W^n_{(i+1)\Delta t} \quad = \quad W^n_{i\Delta t} + \sqrt{\Delta t}\, Z_i$$

for $i \in \mathbb{N}$ and join $W^n_{i\Delta t}$ and $W^n_{(i+1)\Delta t}$ using linear interpolation; see Figure 8 for examples.

```
import numpy as np

class BrownianMotion:
    """
    Defines and simulates a standard Brownian motion
    """
    def __init__(self, T, Nt):
        self.T = T
        self.Nt = Nt
        self.timesteps = np.linspace(0, self.T, num = (Nt+1))

    def simulate(self, nsims = 1):
        x = np.zeros((self.Nt+1, nsims))
        x[0,:] = 0.
        dt = self.T/(self.Nt)
        errs = np.random.randn(self.Nt, nsims)
        for t in range(self.Nt):
            x[t + 1,:] = x[t,:] + np.sqrt(dt) * errs[t,:]
        return x
```

Using the code above we can simulate the paths in Figure 8 as follows:

```
BM_1 = utils.BrownianMotion(T = 1, Nt = 5)
BM_2 = utils.BrownianMotion(T = 1, Nt = 50)
BM_3 = utils.BrownianMotion(T = 1, Nt = 500)

sims_1 = BM_1.simulate(nsims = 3)
time_1 = BM_1.timesteps

sims_2 = BM_2.simulate(nsims = 3)
time_2 = BM_2.timesteps

sims_3 = BM_3.simulate(nsims = 3)
time_3 = BM_3.timesteps
```

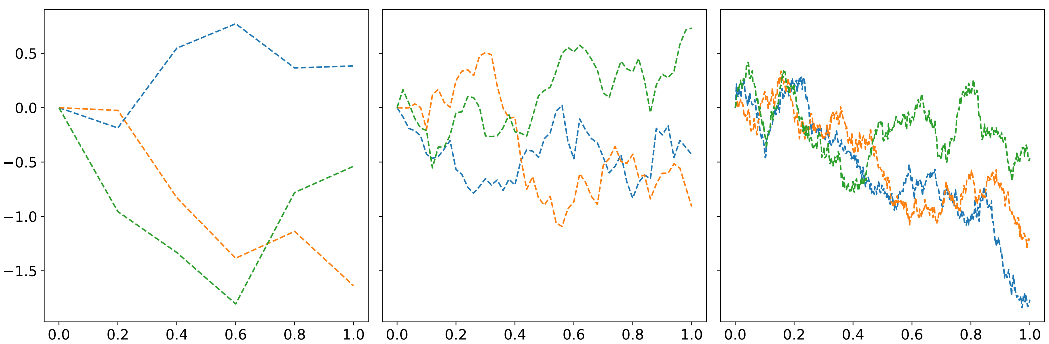

Figure 8: Euler scheme approximation for Brownian motion with large ($\Delta t = 1/5$), medium ($\Delta t = 1/50$), and small ($\Delta t = 1/500$) step size – the randomness is different in each of the plots.

The W_t^n process starts at zero, has independent increments (more precisely the discrete time process $X_i = W_{i\Delta t}^n$ has independent increments) and we see that $W_{(i+1)\Delta t} - W_{i\Delta t} \sim N(0, \Delta t)$, which is consistent with the third property of Brownian motion that $W_t - W_s \sim N(0, t-s)$ (recall that for *any* random variable X, we have that $\text{Var}(aX) = a^2\text{Var}(X)$).

Remark 3.4 *This procedure is known as the* ***Euler method****. If we then join the points*

$$(W_{\Delta t}^n, W_{2\Delta t}^n, W_{3\Delta t}^n, \dots) \tag{89}$$

with straight lines, this methods gives a piecewise linear approximation W_t^n to a true Brownian motion.

Using a deeper result called Donsker's theorem, it can be proved that this construction tends to a Brownian motion as the step size $\Delta t \to 0$, i.e. as $n \to \infty$ (recall that $\Delta t = \frac{1}{n}$). More precisely, for any bounded continuous function $f(x_1, ..., x_n)$ and $0 = t_0 < t_1 < t_2 < ... < t_k$, we have that

$$\lim_{n\to\infty} \mathbb{E}(f(W_{t_1}^n, ..., W_{t_k}^n)) = \mathbb{E}(f(X_1, ..., X_k))$$

where $X = (X_1, .., X_k)$ is a vector of k Normal random variables with $\mathbb{E}(X_i X_j) = R(t_i, t_j)$, where $R(.,.)$ is the covariance function computed above in (78).

Brownian motion can also be obtained by scaled versions of a symmetric random walk. Recall Example 2.11 and take $p = 0.5$; you can think of this as the symmetric coin tosses case. Turns out that by speeding up time and scaling the process one obtains a Brownian motion in the limit. More precisely, if we define

$$W_t^{(n)} = \frac{1}{\sqrt{n}} R_{tn} \,, \tag{90}$$

then, the process

$$\left(W_t^{(n)}\right)_{t\geq 0}$$

converges to a Brownian motion as $n \to \infty$; see Section 3.2 in Shreve (2004) for more details. This result allows to connect what you learnt in 388 with what you will learn in 338.

```
class RandomWalk:
    """
    Defines and simulates a Random Walk
    """
    def __init__(self, p, T, Nt):
        self.p = p
        self.T = T
        self.Nt = Nt
        self.timesteps = np.linspace(0, self.T, num = (Nt+1))
        assert self.Nt>self.T, 'The number of steps should be greater than the
        ↪  terminal time'

    def simulate(self, nsims = 1):
        x = np.zeros((self.Nt+1, nsims))
        x[0,:] = 0.
        errs = (np.random.rand(self.Nt, nsims) <= self.p) * 2. - 1.
        for t in range(self.Nt):
            if floor(self.timesteps[t+1]) > floor(self.timesteps[t]):
                x[t + 1,:] = x[t,:] + errs[t,:]
            else:
                x[t + 1,:] = x[t,:]
        return x
```

Using the above class we simulate the results in Figure 9 as follows.

```
n1 = 10.
n2 = 100.
n3 = 1000.
T = 1.

# The idea is to simulate a standard Random Walk in [0, n*T] and then scale it to
↪  create the approximation between [0,1]
RW_1 = utils.RandomWalk(p = 0.5, T = T*n1, Nt = int(T*n1*2))
RW_2 = utils.RandomWalk(p = 0.5, T = T*n2, Nt = int(T*n2*2))
```

```
RW_3 = utils.RandomWalk(p = 0.5, T = T*n3, Nt = int(T*n3*2))

time1 = RW_1.timesteps
time2 = RW_2.timesteps
time3 = RW_3.timesteps

sim1 = RW_1.simulate(nsims = 3)
sim2 = RW_2.simulate(nsims = 3)
sim3 = RW_3.simulate(nsims = 3)
```

Figure 9 shows three simulations for $\left(W_t^{(n)}\right)_{t\in[0,1]}$ when $n \in \{10, 100, 1000\}$ – see Jupyter notebooks for details if you are curious about the simulations.

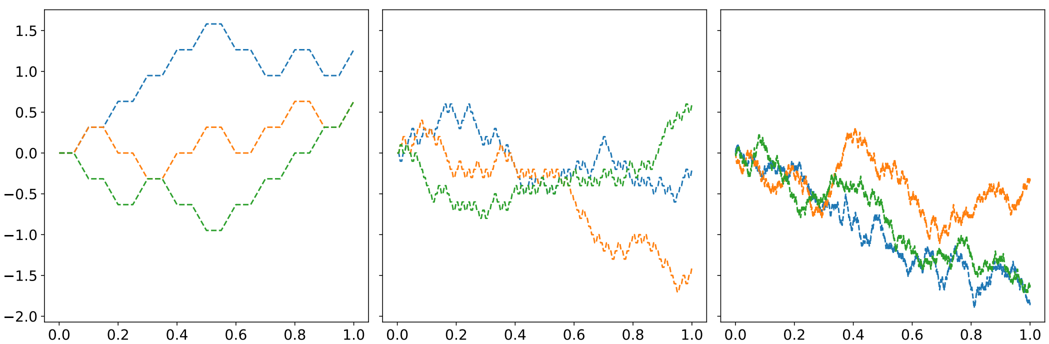

Figure 9: Three simulations of process in (90) for $n = 10$ in the left panel, $n = 100$ in the middle panel, and $n = 1000$ in right panel.

3.4 Quadratic variation of Brownian motion

A partition of the time interval $[0, t]$ is a set of the form $\Pi = \{t_0 = 0 < t_1 < ... < t_n = t\}$, and we define the size of the partition to be

$$\|\Pi\| \quad = \quad \max_{0\le i\le n-1} (t_{i+1} - t_i)$$

i.e. equal to the largest interval of the partition. The **quadratic variation** of a random process X over a fixed time interval $[0, t]$ is then defined as

$$[X, X]_t \quad = \quad \lim_{\|\Pi\|\to 0} \sum_{i=0}^{n-1} (X_{t_{i+1}} - X_{t_i})^2$$

if this limit exists and does not depend on the choice of the sequence of partitions Π. In general the quadratic variation $[X, X]_t$ of a process X is a random process, but we will see that for Brownian motion W, $[W, W]_t = t$ a.s.

Let W denote a standard Brownian motion. We define

$$[W,W]_t^n = \sum_{i=0}^{n-1} (W_{t_{i+1}} - W_{t_i})^2 \tag{91}$$

to be the sampled quadratic variation, for a single partition Π, where $n = n(\Pi)$ denotes the number of partitions in Π.

Proposition 3.10 *The following holds true*

$$\begin{aligned} \mathbb{E}([W,W]_t^n) &= t \\ \mathrm{Var}([W,W]_t^n) = \mathbb{E}(([W,W]_t^n - t)^2) &\to 0 \end{aligned}$$

as $n \to \infty$.

Proof. We first note that

$$\mathbb{E}\left[[W,W]_t^n\right] = \mathbb{E}\left[\sum_{i=0}^{n-1}(W_{t_{i+1}} - W_{t_i})^2\right] = \sum_{i=0}^{n-1}\mathbb{E}[(W_{t_{i+1}} - W_{t_i})^2] = \sum_{i=0}^{n-1}(t_{i+1} - t_i) = t$$

since $W_{t_{i+1}} - W_{t_i} \sim N(0, t_{i+1} - t_i)$. Thus the expected sampled quadratic variation is independent of partition under consideration, and trivially $\lim_{n\to\infty} \mathbb{E}([W,W]_t^n) = t$, since the expectation here does not depend on n. Moreover

$$\begin{aligned}
\mathrm{Var}([W,W]_t^n) &= \mathbb{E}\left[([W,W]_t^n - t)^2\right] = \mathbb{E}\left[\left(\sum_{i=0}^{n-1}(W_{t_{i+1}} - W_{t_i})^2 - t\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{i=0}^{n-1}\left[(W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1} - t_i)\right]\right)^2\right] \\
&\quad \text{(all cross products when squaring the above have expectation zero)} \\
&= \sum_{i=0}^{n-1}\mathbb{E}\left[\left((W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1} - t_i)\right)^2\right] \\
&= \sum_{i=0}^{n-1}\mathbb{E}\left[(W_{t_{i+1}} - W_{t_i})^4 - 2(t_{i+1} - t_i)(W_{t_{i+1}} - W_{t_i})^2 + (t_{i+1} - t_i)^2\right] \\
&= \sum_{i=0}^{n-1} 3(t_{i+1} - t_i)^2 - 2(t_{i+1} - t_i)^2 + (t_{i+1} - t_i)^2 \\
&\quad \text{(using that } W_{t_{i+1}} - W_{t_i} \sim \sqrt{t_{i+1} - t_i}\, Z \text{ and } \mathbb{E}(Z^4) = 3 \text{ where } Z \sim N(0,1)) \\
&= 2\sum_{i=0}^{n-1}(t_{i+1} - t_i)^2 \\
&\leq 2\sum_{i=0}^{n-1}(t_{i+1} - t_i)\,\|\Pi\| \\
&= 2\|\Pi\| t
\end{aligned}$$

which clearly tends to zero if $\|\Pi\| \to 0$. ■

Definition 3.2 *The following are important notions of convergence of random variables:*

(i) A sequence of random variables X_n is said to converge to a random variable X in L^2 if $\mathbb{E}\left[(X - X_n)^2\right] \to 0$.

(ii) A sequence of random variables X_n is said to converge to a random variable X in probability if $\lim_{n\to\infty} \mathbb{P}(|X - X_n| > K) \to 0$ for all $K > 0$.

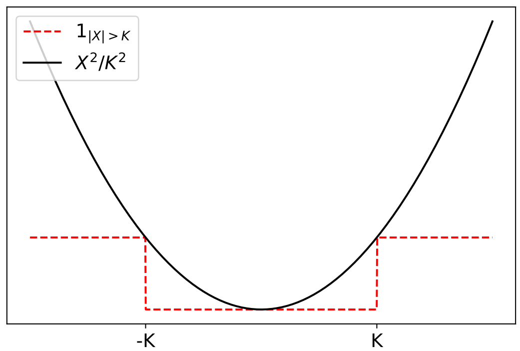

Figure 10: Graph to demonstrate the Chebychev inequality: the black line is $\frac{1}{K^2}x^2$, and the red dotted line is $1_{|x|>K}$. We see that the black line is always above the grey line. We then replace a fixed value of x with a random variable X, and take expectations to obtain the Chebychev inequality.

Proposition 3.11 *The quadratic variation of $(W_t)_{t\geq 0}$ converges to t in L^2.*

Proof. Recall that Proposition 3.10 shows that

$$\mathbb{E}\left[([W,W]_t^n - t)^2\right] \to 0$$

so we see that $[W,W]_t^n$ converges to t in L^2 where here $X_n = [W,W]_t^n$ and $X = t$. ■

Proposition 3.12 *The quadratic variation of $(W_t)_{t\geq 0}$ satisfies that $[W,W]_t^n \to t$ in probability.*

Proof. Using the Chebychev inequality (see Figure 10): for any random variable X:

$$1_{|X|>K} \quad \leq \quad \frac{1}{K^2}X^2$$

so we see that

$$\mathbb{P}(|X| > K) = \mathbb{E}\left[1_{|X|>K}\right] \quad \leq \quad \frac{1}{K^2}\mathbb{E}\left[X^2\right]$$

this implies that

$$\mathbb{P}(|[W,W]_t^n - t| > K) \leq \frac{1}{K^2}\mathbb{E}\left[([W,W]_t^n - t)^2\right]$$

and thus

$$\lim_{\|\Pi\|\to 0} \mathbb{P}(|[W,W]_t^n - t| > K) \leq \lim_{\|\Pi\|\to 0} \frac{1}{K^2}\mathbb{E}\left[([W,W]_t^n - t)^2\right] \quad = \quad 0$$

for all $K > 0$, since we have just shown that

$$\lim_{\|\Pi\|\to 0} \mathbb{E}\left[\left([W,W]_t^{n(\Pi)} - t\right)^2\right] = 0\,. \tag{92}$$

■

Remark 3.5 *See the Jupyter notebook for this section where this is confirmed numerically (note convergence is slow).*

Remark 3.6 *The a.s. convergence (by this we mean that $X_n \to X$ with probability one) implies convergence in probability but convergence in probability does not always imply a.s. convergence.*

4 Stochastic integrals with respect to Brownian motion

Let W be a standard Brownian motion as in the previous section. In this chapter we define a stochastic integral for a large class of integrands with respect to Brownian motion. The main challenge is to overcome the fact that the Brownian motion is not of bounded variation which means that

$$\lim_{\|\Pi_n\|\to 0} \sum_{i=0}^{n-1} |W_{t_{i+1}^n} - W_{t_i^n}| = \infty \,, \tag{93}$$

where Π_n denotes a sequence of partitions with $\|\Pi_n\| \to 0$ as before. This prevents us to apply the integration theory from calculus in a pathwise sense. A way to circumvent this difficulty was introduced by K. Itô in the 1940s by defining the stochastic integral as a limit in L^2.

More precisely, we want to make sense of an integral of the form $\int_0^t \alpha_s dW_s$ where α is a stochastic process satisfying some integrability constraint. It is natural to only consider adapted process α; this means that the value of α_t for each t depends only on the history of W up to time t, i.e. α cannot see into the future, which will be natural in a financial context later when we consider dynamic trading strategies under the famous Black-Scholes model. For example, the process

$$\alpha_t = \int_0^t f(W_s) ds \,, \tag{94}$$

is an adapted process but $\alpha_t = W_{t+\delta}$ is not if $\delta > 0$. An integral of the form

$$\int_0^t \alpha_s dW_s \tag{95}$$

is known as a stochastic integral. Next, Section 4.1 motivates why do we need a new type of integration and Section 4.2 constructs the stochastic integral.

4.1 Why do we need a new type of integration?

The standard approach to define an integral for a deterministic function $f : [a, b] \to \mathbb{R}$ is the Riemann integral which is defined by

$$\int_a^b f(s) ds := \lim_{\|\Pi_n\|\to 0} \sum_{i=0}^{n-1} f(\xi_i) \, (t_{i+1} - t_i) \,, \tag{96}$$

and $\Pi_n = \{a = t_0 < t_1 < \cdots < t_n = b\}$ is a partition of $[a, b]$ and $\xi_i \in [t_i, t_{i+1}]$. Of course, this definition makes only sense if the right hand side converges for every sequence of partitions and choices of ξ. In this case, the function f is called Riemann integrable and the left hand side is the Riemann integral. One can show that at least every continuous function $f : [a, b] \to \mathbb{R}$ is Riemann integrable. The Riemann integral can be generalised to the Riemann-Stieltjes integral which assigns various weights to the integrand f. The weights are described by a function $g : [a, b] \to \mathbb{R}$. Then the Riemann-Stieltjes integral is defined by

$$\int_a^b f(s) g(ds) := \lim_{\|\Pi_n\|\to 0} \sum_{i=0}^{n-1} f(\xi_i) \, (g(t_{i+1}) - g(t_i)) \,, \tag{97}$$

where as before, the definition makes only sense if the right hand side converges for every sequence of partitions and choices of ξ. In this situation, the function f is called Riemann-Stieltjes integrable with respect to g and the unique limit is called the Riemann-Stieltjes integral of f with respect to g. One can show that at least every continuous function $f : [a, b] \to \mathbb{R}$ is Riemann-Stieltjes integrable with respect to every function g of finite variation.

In general, the integrand f can level out irregularities of the integrator g such that the limit in (97) exists and vice versa. However, if we ask for the largest class of integrators g such that at least every continuous function is Riemann-Stieltjes integrable with respect to g, then g must be of finite variation (see Theorem 56 in Protter (2005)). This result rules out the possibility to define the stochastic integral of a stochastic process $(\alpha_t)_{t\geq 0}$ with respect to a Brownian motion pathwise. That is, for fixed $\omega \in \Omega$ define an integral of the function $t \to f_\omega(t) := \alpha_t(\omega)$ with respect to the function $t \to g_\omega(t) := W_t(\omega)$, which turns out to be impossible since the function g_ω is of infinite variation for $\mathbb{P} - a.a.\ \omega \in \Omega$.

4.2 The construction

Let $(W_t)_{t\in[0,T]}$ be a Brownian motion and let us work with $\mathcal{F}_t = \sigma(W_s : 0 \leq s \leq t)$.

Definition 4.1 *A stochastic process $(\alpha_t)_{t\in[0,T]}$ is called a simple stochastic process if it is of the form*

$$\alpha_t = \sum_{i=1}^{N} \alpha_i 1_{t\in(t_{i-1},t_i]}\,, \tag{98}$$

where α_i is random but only depends on the history of W up to time t_i, and $0 = t_0 < t_1 < \cdots < t_N = T$.

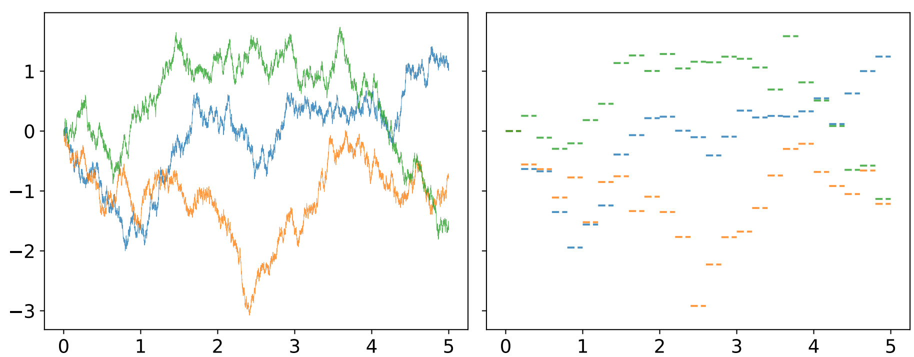

Figure 11: Example of a simple stochastic process α_t. Here, we take $t_0 = 0$, $t_N = 5$, equal spacing $\Delta t = 0.2$, and $\alpha_t = W_{t_{i-1}}$ for $t \in (t_{i-1}, t_i]$. The left panel has three sample paths of a standard Brownian motion, and the right panel shows the simple stochastic process just defined.

Definition 4.2 *Let $(\alpha_t)_{t\in[0,T]}$ be a simple stochastic process, we define the stochastic integral of α_t*

with respect to W_t *by*

$$I(\alpha) := \int_0^T \alpha_s dW_s \quad = \quad \sum_{i=1}^{N} \alpha_i \left(W_{t_i} - W_{t_{i-1}}\right).$$

To generalize the stochastic integral to more general (non-simple) process α_t, we approximate the process to arbitrary accuracy with a simple process and use arguments involving L^2-convergence. More precisely, we can show that the integral $I(\alpha)$ satisfies the so-called Ito's isometry for simple stochastic processes, that is

$$\mathbb{E}\left[(I(\alpha))^2\right] = \int_0^T \mathbb{E}\left[(\alpha_s)^2\right] ds\,. \tag{99}$$

It can also be shown that the space $L^2_{\mathbb{P}}(\Omega)$ defined by the random variables that satisfy

$$\mathbb{E}[X^2] < \infty \tag{100}$$

is a Hilbert space with norm and inner product given by

$$\|X\|_{L^2_{\mathbb{P}}} = \left(\mathbb{E}\left[X^2\right]\right)^{1/2}\,, \qquad \langle X, Y\rangle_{L^2_{\mathbb{P}}} = \mathbb{E}[X\,Y]\,, \tag{101}$$

similarly, the space of simple stochastic processes that satisfy

$$\int_0^T \mathbb{E}\left[\alpha_s^2\right] ds < \infty\,, \tag{102}$$

is also a Hilbert space with norm and inner product given by

$$\|\alpha\|_{L^2_{ds\otimes\mathbb{P}}} = \left(\int_0^T \mathbb{E}\left[\alpha_s^2\right] ds\right)^{1/2}\,, \qquad \langle \alpha, \beta\rangle_{L^2_{ds\otimes\mathbb{P}}} = \int_0^T \mathbb{E}[\alpha_s\,\beta_s] ds\,. \tag{103}$$

Lastly, it turns out that for any adapted stochastic process Ψ satisfying (102) there is a sequence $\{H_n\}_{n\in\mathbb{N}}$ of simple stochastic processes converging to Ψ

$$\lim_{n\to\infty} \|H_n - \Psi\|^2_{L^2_{ds\otimes\mathbb{P}}} = 0\,, \tag{104}$$

then, by the triangle inequality of norms it follows that

$$\lim_{m,n\to\infty} \|H_m - H_n\|^2_{L^2_{ds\otimes\mathbb{P}}} = 0\,, \tag{105}$$

and by Ito's isometry for simple stochastic processes we have that

$$\|H_m - H_n\|^2_{L^2_{ds\otimes\mathbb{P}}} = \|I(H_m) - I(H_n)\|^2_{L^2_{\mathbb{P}}}\,, \tag{106}$$

for all $m, n \in \mathbb{N}$, thus, we obtain that

$$\lim_{m,n\to\infty} \|I(H_m) - I(H_n)\|^2_{L^2_{\mathbb{P}}} = 0\,. \tag{107}$$

In other words, $\{I(H_n)\}_{n\in\mathbb{N}}$ is a Cauchy sequence in the Hilbert space $L^2_{\mathbb{P}}$ and it is well known that each Cauchy sequence converges to an element in the Hilbert space. This means that there is $Y \in L^2_{\mathbb{P}}$ such that

$$\lim_{n\to\infty} \|I(H_n) - Y\|^2_{L^2_{\mathbb{P}}} = 0\,. \tag{108}$$

We call Y the stochastic integral of Ψ and define

$$I(\Psi) := \int_0^T \Psi_s dW_s := \lim_{n\to\infty} \int_0^T H_n(s) dW_s \qquad \text{in } L^2_{\mathbb{P}}\,. \tag{109}$$

4.3 Stochastic differential equations

We now consider a stochastic differential equation (SDE) which we can write informally as

$$dX_t \;=\; b(X_t)dt \,+\, \sigma(X_t)dW_t\,, \qquad X_0 = x_0 \tag{110}$$

and enquire, can we assign a rigorous meaning to this equation? Note that if $b(x) = 0$ and $\sigma(x) = 1$, then $dX_t = dW_t$, so $X_t = x_0 + W_t$.

Definition 4.3 *A strong solution to the SDE (110) with $X_0 = x_0$ is a process X_t with a continuous sample path which satisfies the integral equation*

$$X_t \;=\; x_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s\,, \tag{111}$$

here $b(\cdot)$ is known as the drift, and $\sigma(\cdot)$ is known as the volatility.

Remark 4.1 *Note if $\sigma \equiv 0$, then (110) reduces to the Ordinary Differential equation (ODE):*

$$\frac{dx(t)}{dt} \;=\; b(x(t))$$

which we can solve be re-writing as $\frac{dx}{b(x)} = dt$ and integrating both sides.

Definition 4.4 *A function $f : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous if $\exists\, K \in \mathbb{R}^+$ such that for $x, y \in \mathbb{R}$*

$$|f(x) - f(y)| \leq K\,|x - y|\,. \tag{112}$$

Proposition 4.1 *If b and σ are Lipschitz continuous, then a strong solution to (110) exists.*

Note that X in (111) is implicitly defined in terms of itself, so how do we construct a solution rigorously? We first simulate the Brownian sample path W. We then start with the constant solution $X_t^{(0)} = x_0$ for all $t \in [0, T]$, feed it into the SDE, and then we keep feeding the answer back into the SDE as follows:

$$\begin{aligned}
X_t^{(1)} &= x_0 + \int_0^t b(X_0^{(0)})ds + \int_0^t \sigma(X_0^{(0)})dW_s\,,\\
X_t^{(2)} &= x_0 + \int_0^t b\left(X_s^{(1)}\right)ds + \int_0^t \sigma\left(X_s^{(1)}\right)dW_s\\
&\;\;\cdots\\
X_t^{(n+1)} &= x_0 + \int_0^t b\left(X_s^{(n)}\right)ds + \int_0^t \sigma\left(X_s^{(n)}\right)dW_s\\
&\;\;\cdots
\end{aligned}$$

With some work we can then show that the sequence of sample paths $\{X_s^{(n)}; 0 \leq s \leq t\}$ converges to some continuous function X_t which we call the solution to (110), more precisely,

$$\lim_{n\to\infty} \max_{0\leq s\leq t} |X_s^n - X_s| \to 0\,, \tag{113}$$

see chapter 5 in Karatzas and Shreve (1991) for details. This is called the Picard iteration method.

On a computer, we typically do not use the Picard method to approximate the SDE numerically, but rather we use an extension of the Euler scheme where we essentially "freeze" the coefficients over each time step as follows:

$$X^n_{t+\Delta t} \quad = \quad X^n_t \; + \; b(X^n_t)\Delta t \; + \; \sigma(X^n_t)\sqrt{\Delta t}\; Z_i \tag{114}$$

where Z_i is a sequence of i.i.d. standard Normals and $\Delta t = \frac{1}{n}$ as before, and under certain conditions we can show that the approximate solution X^n_t tends to true solution X_t to the SDE as $n \to \infty$ (i.e. as the step size $\Delta t = \frac{1}{n} \to 0$).

4.4 Ito's lemma

Remark 4.2 *We use the following notation for partial derivatives:*

$$\partial_x f \equiv f_x \,,$$
$$\partial_x \partial_y f \equiv \partial_{xy} f \equiv f_{xy} \,.$$

Theorem 4.2 *Let $(X_t)_{t\in[0,T]}$ be a process satisfying the SDE in (110) and let $f(t,x)$ be a function which is once differentiable in t and twice differentiable in x.*[3] *Then*

$$\begin{aligned} f(t,X_t) &= f(0,X_0) + \int_0^t f_x(s,X_s) dX_s \; + \; \int_0^t \left[f_t(s,X_s) + \frac{1}{2} f_{xx}(s,X_s)\,\sigma(X_s)^2 \right] ds \\ &= f(0,X_0) + \int_0^t f_x(s,X_s)\,[b(X_s)ds + \sigma(X_s)dW_s] \; + \; \int_0^t \left[f_t(s,X_s) + \frac{1}{2} f_{xx}(s,X_s)\sigma(X_s)^2 \right] ds \,. \end{aligned} \tag{115}$$

Proof. **(Idea of proof)** For simplicity assume $f(t,x) \equiv f(x)$ and $X_t = W_t$. Given that W satisfies the trivial SDE

$$dX_t = 0dt \; + \; 1dW_t \,, \quad X_0 = 0 \,,$$

we wish to show that for f twice differentiable

$$f(W_t) = f(W_0) + \int_0^t f_x(W_s) dW_s + \frac{1}{2} \int_0^t f_{xx}(W_s)\, ds \,. \tag{116}$$

Consider a partition $0 = t_0 < \cdots < t_n = t$ of $[0,t]$. Then, by Taylor's theorem,

$$f(W_t) - f(W_0) = \sum_{i=0}^{n-1} \big(f(W_{t_{i+1}}) - f(W_{t_i}) \big) \tag{117}$$

$$= \sum_{i=0}^{n-1} f_x(W_{t_i}) \left(W_{t_{i+1}} - W_{t_i} \right) \tag{118}$$

$$+ \frac{1}{2} \sum_{i=0}^{n-1} f_{xx} \left(W_{t_i} + \theta_i \, \left(W_{t_{i+1}} - W_{t_i} \right) \right) \left(W_{t_{i+1}} - W_{t_i} \right)^2 \,, \tag{119}$$

[3] Mathematically we write this as $f \in \mathcal{C}^{1,2}([0,T] \times \mathbb{R})$.

for $\theta_i \in [-1,1]$. As the partition becomes finer, it can be shown that (118) approximates the stochastic integral

$$\int_0^t f_x(W_s)dW_s\,, \tag{120}$$

and that (119) approximates

$$\frac{1}{2}\int_0^t f_{xx}(W_s)\,ds\,. \tag{121}$$

This last assertion is consequence of what we studied on the quadratic variation of Brownian motion. ■

As shorthand, we write (115) in the differential form as

$$df(t,X_t) \;=\; f_t(t,X_t)dt \,+\, f_x(t,X_t)dX_t \,+\, \frac{1}{2}f_{xx}(t,X_t)\sigma(X_t)^2dt\,.$$

Note that we get a surprising additional second order term here (the term involving f_{xx}) which we do not get if X_t is just a non-random function which differentiable in x and in t), and this extra terms is the main difference between ordinary calculus and stochastic calculus.

Corollary 4.3 *Let $f(t,x)$ be a function which is once differentiable in t and twice differentiable in x and define $Y_t = f(t,W_t)$ where W is a Brownian motion, then*

$$dY_t = \left(f_t(t,W_t) + \frac{1}{2}\,f_{xx}(t,W_t)\right)dt + f_x(t,W_t)\,dW_t\,. \tag{122}$$

4.5 Important examples

Here, we provide important examples of how one uses Ito's lemma.

(i) Let $Y_t = W_t^2 - t$. Find the SDE satisfied by Y_t.

Let $f(t,x) = x^2 - t$ so $f(t,W_t) = W_t^2 - t$. Then $f_x(t,x) = 2x$, $f_{xx}(t,x) = 2$ and $f_t(t,x) = -1$. Thus from Ito's lemma we have

$$\begin{aligned} dY_t = df(t,W_t) &= f_t(t,W_t)dt + f_x(t,W_t)dW_t + \frac{1}{2}f_{xx}(t,W_t)dt \\ &= -1\,dt \,+\, 2\,W_t\,dW_t \,+\, \frac{1}{2}\,2\,dt \\ &= 2\,W_t\,dW_t\,. \end{aligned}$$

Writing this in integrated form we have $f(t,W_t) - f(0,W_0) = W_t^2 - t = \int_0^t 2\,W_s\,dW_s$, and we can easily verify that $M_t = W_t^2 - t$ satisfies $\mathbb{E}[Y_t|Y_s] = Y_s$.

(ii) Let $S_t = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}$. Find the SDE satisfied by S_t.

Let $f(t,x) = S_0 e^{(\mu-\frac{1}{2}\sigma^2)t+\sigma x}$, so $S_t := f(t,W_t) = S_0 e^{(\mu-\frac{1}{2}\sigma^2)t+\sigma W_t}$. This is the famous **Black-Scholes** model for a share price process. Then $f_x(t,x) = \sigma f(t,x)$, $f_{xx}(t,x) = \sigma^2 f(t,x)$

and $f_t(t,x) = (\mu - \frac{1}{2}\sigma^2)f(t,x)$. Thus from Ito's lemma we have

$$
\begin{aligned}
dS_t &= f_t(t,W_t)dt + f_x(t,W_t)\,dW_t + \frac{1}{2}f_{xx}(t,W_t)\,dt \\
&= \left(\mu - \frac{1}{2}\sigma^2\right) S_t dt \,+\, \sigma S_t dW_t \,+\, \frac{1}{2}\sigma^2 S_t dt \\
&= \mu\, S_t\, dt \,+\, \sigma\, S_t\, dW_t \qquad (123) \\
&= b(S_t)\, dt \,+\, \tilde{\sigma}(S_t)\, dW_t
\end{aligned}
$$

where $b(S) = \mu\, S$ and $\tilde{\sigma}(S) = \sigma\, S$. Note that b and σ are both linear functions of S, so this process is known as a **geometric Brownian motion** (GBM).

The parameter μ describes the overall trend of the process, i.e. its tendency to go up or down in the long run (it can be shown using the mgf of a Normal distribution that $\mathbb{E}[S_t] = S_0 e^{\mu t}$, and σ is the volatility which controls the variability of the share price. Note that S_t is always positive. Figure 12 illustrates the influence of μ and σ with simulations.

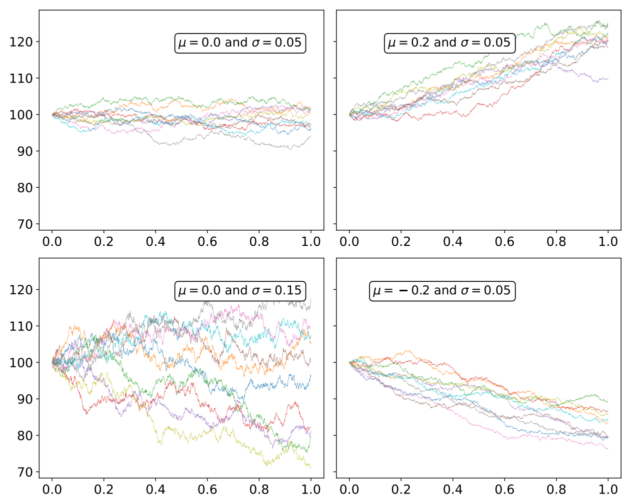

Figure 12: Ten simulations of a GBM for different configurations of the parameters μ and σ. The left panels show the influence of σ and the right panels show the influence of μ.

(iii) **pth power of share price for Black-Scholes model**. Assume that $dS_t = \sigma\, S_t\, dW_t$ and let $M_t = S_t^p$ for $p > 1$. Apply Ito's lemma to M_t, and write dM_t as an SDE in terms of M_t itself.

Set $M_t = f(t, S_t)$, where $f(t,S) = S^p$. Applying Ito's lemma to $dS_t = \sigma\, S_t\, dW_t$, we have that

$f_t(t,S)=0,\ f_S(t,S)=p\,S^{p-1}, f_{SS}(t,S)=p\,(p-1)\,S^{p-2}$ and

$$\begin{aligned} dM_t \;=\; df(t,S_t) &= f_t(t,S_t)\,dt \,+\, f_S(t,S_t)dS_t \,+\, \frac{1}{2}f_{SS}(t,S_t)\,\sigma^2\,S_t^2\,dt \\ &= p\,S_t^{p-1}\,dS_t \,+\, \frac{1}{2}p\,(p-1)\,S_t^{p-2}\,S_t^2\,\sigma^2\,dt \\ &= p\,S_t^{p-1}\,S_t\,\sigma\,dW_t \,+\, \frac{1}{2}\,p\,(p-1)\,S_t^p\,\sigma^2\,dt \\ &= \frac{1}{2}\,p\,(p-1)\,\sigma^2\,M_t\,dt \,+\, p\,\sigma\,M_t\,dW_t \\ &= M_t\left(\frac{1}{2}\,p\,(p-1)\,\sigma^2\,dt \,+\, p\,\sigma\,dW_t\right)\,, \end{aligned}$$

and we see that the drift and volatility of M_t are linear functions of M_t. Thus, M is a geometric Brownian motion similar to S in (123).

(iv) **The Ornstein-Uhlenbeck process**. Consider the Ornstein-Uhlenbeck (OU) process which satisfies

$$dY_t \;=\; -\kappa\,Y_t\,dt+\sigma\,dW_t \tag{124}$$

for $\kappa,\sigma>0$. Using $Z_t=e^{\kappa t}Y_t$ find a solution to Y_t.

The drift term means the process tends to "mean-revert" back to zero if Y goes too far away from zero, and more generally we can let $dY_t=\kappa(\theta-Y_t)dt+\sigma dW_t$, and in this case Y will mean-revert around the level θ. For this example we keep $\theta=0$ for simplcitiy, and let $Z_t=e^{\kappa t}Y_t=f(t,Y_t)$ where $f(t,y)=e^{\kappa t}y$. Then $f_t=\kappa\,f$, $f_y=e^{\kappa t}$ and $f_{yy}=0$, and applying Ito's lemma to Y_t directly we see that

$$\begin{aligned} dZ_t \;=\; f_tdt+f_ydY_t+\frac{1}{2}f_{yy}\sigma^2dt &= \kappa e^{\kappa t}Y_tdt \,+\, e^{\kappa t}dY_t+0 \\ &= \kappa e^{\kappa t}Y_tdt \,+\, e^{\kappa t}(-\kappa Y_tdt+\sigma dW_t) \\ &= \sigma e^{\kappa t}dW_t\,. \end{aligned}$$

Integrating this equation, we see that $Z_t=Z_0+\int_0^t\sigma e^{\kappa s}dW_s$. Multiplying by $e^{-\kappa t}$ and noting that $Z_0=Y_0$, we obtain

$$Y_t \;=\; e^{-\kappa t}Z_0 \,+\, e^{-\kappa t}\int_0^t\sigma e^{\kappa s}dW_s \;=\; e^{-\kappa t}Y_0 \,+\, \int_0^t\sigma e^{-\kappa(t-s)}dW_s\,.$$

This is the solution to the SDE in (124). It turns out that for any non-random continuous function ϕ we have that

$$\int_0^t\phi(s)dW_s\sim N\left(0,\int_0^t\phi(s)^2d\,s\right)\,. \tag{125}$$

The proof of this result is not relevant for the course. A straightforward proof can be put forward for when σ is piecewise constant. In this case, setting $\phi(s)=\sigma e^{-\kappa(t-s)}$, we find that

$$\begin{aligned} Y_t &\sim e^{-\kappa t}Y_0 \,+\, N\left(0,\int_0^t\phi(s)^2ds\right) \\ &= e^{-\kappa t}Y_0 \,+\, N\left(0,\,\sigma^2\frac{1-e^{-2\kappa t}}{2\kappa}\right) \end{aligned}$$

From this we see that $Y_\infty := \lim_{t\to\infty} Y_t \sim N(0, \frac{\sigma^2}{2\kappa})$. We call this the **stationary distribution** of Y. The OU process is often used in practice to model volatility or an interest rate process (this is the well known **Vasicek model**, see FM07 for more details).

(v) **The Black-Scholes model, and foreign exchange rates**. Set $Y_t = 1/S_t$ where S_t satisfies the Black-Scholes SDE $dS_t = \sigma S_t dW_t$ with $\mu = 0$. Find the SDE satisfied by Y_t. Is Y_t a geometric Brownian motion?

Let $f(S) = 1/S$. It follows that $f_S = -\frac{1}{S^2}, f_{SS} = \frac{2}{S^3}$ and therefore

$$dY_t \quad = \quad -\frac{1}{S_t^2}dS_t + \frac{1}{2}\frac{2}{S_t^3}S_t^2\sigma^2 dt \quad = \quad Y_t\left(\sigma^2 dt \,-\, \sigma dW_t\right)\,.$$

If S_t is e.g. the GBP/USD Exchange rate i.e. the cost of a pound in dollars, then $Y_t = 1/S_t$ is the USD/GBP exchange rate, i.e. the cost of a dollar in pounds. We see that Y_t is a GBM.

5 Continuous-time models

5.1 The Black-Scholes model

As mentioned briefly before, the Black-Scholes model for a share price process is defined as

$$S_t \quad = \quad S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t} \tag{126}$$

where W is a standard Brownian motion, and from Ito's lemma, we have seen that S_t satisfies the SDE

$$dS_t \quad = \quad S_t(\mu dt \, + \, \sigma dW_t)$$

which is known as geometric Brownian motion. Here, μ is the **drift** of the process which describes the overall upward/downward trend of S, and σ is the **volatility**, which describes the variability of S – see Figure 12.

5.2 The terminal share price distribution

Proposition 5.1 *Let $(S_t)_{t \geq 0}$ be a geometric Brownian motion. Then,*

$$\log S_t \sim N\left(\log S_0 + \left(\mu - \frac{1}{2}\sigma^2\right)t, \sigma^2 t\right) . \tag{127}$$

Proof. Re-arranging (126) we see that

$$\log S_t \, - \, \log S_0 \quad = \quad \log \frac{S_t}{S_0} \quad = \quad \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t \, .$$

Thus we can calculate the distribution of $\log \frac{S_t}{S_0}$ as

$$\log \frac{S_t}{S_0} \sim N\left(\left(\mu - \frac{1}{2}\sigma^2\right)t, \sigma^2 t\right) \tag{128}$$

or equivalently

$$\log S_t \sim N\left(\log S_0 + \left(\mu - \frac{1}{2}\sigma^2\right)t, \sigma^2 t\right) , \tag{129}$$

where we have used that $W_t \sim N(0, t)$ and for any random variable X, $\mathrm{Var}(aX) = a^2 \mathrm{Var}(X)$. ∎

For some fixed $T > 0$, we can then compute $\mathbb{P}(S_T > K)$ as follows:

$$\begin{aligned} \mathbb{P}(S_T > K) \quad &= \quad \mathbb{P}\left(\log \frac{S_T}{S_0} > \log \frac{K}{S_0}\right) \\ &= \quad \mathbb{P}\left(\frac{\log \frac{S_T}{S_0} - (\mu - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} > \frac{\log \frac{K}{S_0} - (\mu - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right) \\ &= \quad \mathbb{P}\left(Z > \frac{\log \frac{K}{S_0} - (\mu - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right) \\ &= \quad \Phi^c(z) \end{aligned} \tag{130}$$

where z in the equation above is given by

$$z = \frac{\log \frac{K}{S_0} - (\mu - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}, \tag{131}$$

and (130) follows because

$$Z = \frac{\log \frac{S_T}{S_0} - (\mu - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} \sim N(0,1), \tag{132}$$

and as usual $\Phi^c(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$. We have used that standard result that $(X-\mu)/\sigma \sim N(0,1)$ if $X \sim N(\mu, \sigma^2)$.

Example 5.1 *Set $S_0 = 1$, $\sigma = 0.1$, $T = 0.25$, and $\mu = 0.05$. Then we have $z = 1.681153596$ and $\mathbb{P}(S_T > K) = \Phi^c(z) = 1 - \Phi(z) = 0.046361687633$.*

We can use `Normsdist(x)` in Excel or `scipy.stats.norm().cdf(x)` in Python to calculate $\Phi(x)$.

Proposition 5.2 *Let $(S_t)_{t\geq 0}$ be a geometric Brownian motion. Then, the density function of S_t denoted by $p_{S_t}(S)$ is given by*

$$p_{S_t}(S) = \frac{1}{S\sigma\sqrt{2\pi t}} \exp\left(- \frac{\left(\log \frac{S}{S_0} - \left(\mu - \frac{1}{2}\sigma^2\right) t\right)^2}{2\sigma^2 t} \right)$$

for $S > 0$.

Proof. Note that

$$\mathbb{P}(S_t \leq S) \quad = \quad \mathbb{P}(\log S_t \leq \log S) \quad = \quad F(\log S)$$

where F is the distribution function of $\log S_t$. Differentiating both sides with respect to S and using the chain rule, we see that the density $p_{S_t}(S)$ of S_t is given by

$$p_{S_t}(S) \quad = \quad \frac{d}{dS}\mathbb{P}(S_t \leq S) \quad = \quad \frac{1}{S}F'(\log S) \quad = \quad \frac{1}{S}p_{X_t}(x)$$

where $x = \log S$ and $p_{X_t}(x)$ is the density of $X_t = \log S_t$ which is given by

$$p_{X_t}(x) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left(- \frac{(x - x_0 - (\mu - \frac{1}{2}\sigma^2)t)^2}{2\sigma^2 t} \right), \tag{133}$$

since $X_t \sim N(X_0 + (\mu - \frac{1}{2}\sigma^2)t, \sigma^2 t)$, and recall that the density of a general $N(\mu_1, \sigma_1^2)$ random variable is given by

$$p(x) \quad = \quad \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$$

so

$$p_{S_t}(S) = \frac{1}{S\sigma\sqrt{2\pi t}} \; \exp\left(- \frac{\left(\log \frac{S}{S_0} - \left(\mu - \frac{1}{2}\sigma^2\right) t\right)^2}{2\sigma^2 t} \right)$$

for $S > 0$. Here S_t has what is known as a **lognormal distribution**. Note the presence of the $\frac{1}{S}$ pre-factor, and this pdf is only defined for $S > 0$ because the share price cannot go negative. The function $p_{S_t}(S)$ is a pdf and thus must integrate to one, i.e., $\int_0^\infty p_{S_t}(S) dS = 1$. ■

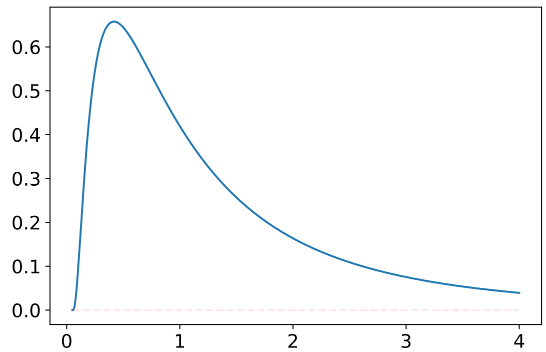

Figure 13: Here we have plotted the share price density $f_{S_t}(S)$ for when $S_0 = 1$, $\mu = 0.05$, $\sigma = 1$ and $t = 1$.

5.3 The Black-Scholes PDE

Let us consider a market with a risk-free bond B_t with $B_0 = 1$ and continuously compounding interest rate $r > 0$; see Appendix 8.1 to build some intuition about continuous compounding. This means that

$$B_t = e^{r\,t}\,, \tag{134}$$

or alternatively,

$$dB_t = r\,B_t\,dt\,. \tag{135}$$

In this market there is a risky asset that follows a GBM, that is

$$dS_t = \mu\,S_t\,dt + \sigma\,S_t\,dW_t\,, \tag{136}$$

with $S_0 > 0$. We consider a contingent claim that pays $f(S_T)$ at time $T > 0$ and some payoff function f. An important example of payoff function f is $f(x) = \max(x - K, 0)$.

Definition 5.1 *A European call option written on the stock $(S_t)_{t\in[0,T]}$ with maturity $T > 0$ and strike $K > 0$ is a derivative that gives the holder the right to buy the security for price K at time T.*

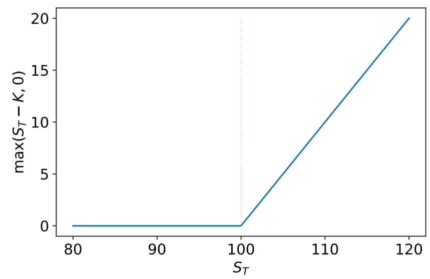

Figure 14: Payoff of a European call option for $K = 100$.

We wish to find what is the price of a derivative that pays $f(S_T)$ at time T. For this, we will set-up a self-financed portfolio at time zero that will have value $f(S_T)$ at time T. Using the no-arbitrage principle we will conclude that the amount needed to set-up such a portfolio will be the price of the derivative.

Definition 5.2 *A trading strategy $(\phi_t, \psi_t)_{t\in[0,T]}$ is a pair of adapted stochastic processes such that ϕ_t is the number of shares the investor holds at time t and ψ_t is the number of risk-free bonds held at time t. Thus, the value of the position of the agent at time t is*

$$V_t = \phi_t\, S_t + \psi_t\, B_t\,. \tag{137}$$

The strategy is called self-financed if

$$dV_t = \phi_t dS_t + \psi_t dB_t\,.^{4} \tag{138}$$

Let $(\phi_t, \psi_t)_{t\in[0,T]}$ be a self-financed trading strategy. Then, the value of the portfolio is that in (137). We want to find a self-financed strategy (ϕ_t, ψ_t) such that $V_T = f(S_T)$. Given that (ϕ_t, ψ_t) is self-financed we have that

$$\begin{aligned} dV_t &= \phi_t dS_t + \psi_t dB_t && (139)\\ &= \phi_t dS_t + \psi_t\, r\, B_t\, dt && (140)\\ &= \phi_t dS_t + r\, (V_t - \phi_t\, S_t)\, dt && (141)\\ &= \phi_t\, (\mu\, S_t\, dt + \sigma\, S_t\, dW_t) + r\, (V_t - \phi_t\, S_t)\, dt\,, && (142)\end{aligned}$$

thus, we have that

$$dV_t = (\phi_t\, \mu\, S_t + (V_t - \phi_t\, S_t)\, r)\, dt + \phi_t\, \sigma\, S_t\, dW_t\,. \tag{143}$$

[4]See Appendix 8.2 to get intuition about the self-financed equation.

Write $V_t = v(t, S_t)$ for $v \in \mathcal{C}^{1,2}([0,T] \times \mathbb{R})$.[5] Then, by Itô's lemma we have that

$$dV_t = \partial_t v(t, S_t)dt + \partial_s v(t, S_t)\ (\mu\, S_t\, dt + \sigma\, S_t\, dW_t) + \frac{1}{2}\, \partial_{ss} v(t, S_t)\, \sigma^2\, S_t^2\, dt \tag{145}$$

$$= \left(\partial_t v(t, S_t) + \partial_s v(t, S_t)\, \mu\, S_t + \frac{1}{2}\partial_{ss} v(t, S_t)\sigma^2\, S_t^2\right) dt + \partial_s v(t, S_t)\, \sigma\, S_t\, dW_t\,. \tag{146}$$

We choose $\phi_t = \partial_s v(t, S_t)$ which is adapted and a function of t and S_t. Then (143) becomes

$$dV_t = (\partial_s v(t, S_t)\, \mu\, S_t + (v(t, S_t) - \partial_s v(t, S_t)\, S_t)\, r)\, dt + \partial_s v(t, S_t)\, \sigma\, S_t\, dW_t\,. \tag{147}$$

By comparing (146) and (147) we see that v should satisfy the following PDE

$$\partial_s v(t, S)\, \mu\, S + (v(t, S) - \partial_s v(t, S)\, S)\, r = \partial_t v(t, S) + \partial_s v(t, S)\, \mu\, S + \frac{1}{2}\partial_{ss} v(t, S)\sigma^2\, S^2\,, \tag{148}$$

and we see that the terms $\partial_s v(t, S)\, \mu\, S$ cancels out and the above PDE reduces to

$$v(t, S)\, r - \partial_s v(t, S)\, S\, r - \partial_t v(t, S) - \frac{1}{2}\partial_{ss} v(t, S)\sigma^2\, S^2 = 0\,. \tag{149}$$

Before giving the main the following result.

Proposition 5.3 *There is a unique solution $v(t, S)$ to the PDE*

$$v(t, S)\, r - \partial_t v(t, S) - \partial_s v(t, S)\, S\, r - \frac{1}{2}\partial_{ss} v(t, S)\sigma^2\, S^2 = 0\,, \tag{150}$$

with terminal boundary condition $v(T, S) = f(S)$.

Using the above proposition we have the following (powerful) result:

Theorem 5.4 *Let $v(t, S)$ be the solution to (150) with terminal condition $f(S)$. The trading strategy*

$$\phi_t = \partial_S v(t, S_t)\,, \qquad \psi_t = (v(t, S_t) - S_t\, \partial_S v(t, S_t))\, e^{-r\,t}\,, \tag{151}$$

is a self-financed trading strategy and

$$V_T = \phi_T\, S_T + \psi_T\, e^{r\,T} = f(S_T)\,. \tag{152}$$

Proof. Let $v(t, S)$ be the solution to (150) with terminal condition $f(S)$ and define

$$\phi_t = \partial_S v(t, S_t)\,, \qquad \psi_t = (v(t, S_t) - S_t\, \partial_S v(t, S_t))\, e^{-r\,t}\,, \tag{153}$$

with $V_t = \phi_t\, S_t + \psi_t\, B_t$. Then,

$$V_t = \phi_t\, S_t + \psi_t\, B_t \tag{154}$$

$$= \partial_S v(t, S_t)\, S_t + (v(t, S_t) - S_t\, \partial_S v(t, S_t))\, e^{-r\,t}\, e^{r\,t} \tag{155}$$

$$= v(t, S_t)\,, \tag{156}$$

[5]Note that this is up to us. Indeed, if we choose a self-financed trading strategy (ϕ_t, ψ_t) where $\phi_t = p(t, S_t)$ and $\psi_t = q(t, S_t)$ for $p, q \in \mathcal{C}^{1,2}([0,T] \times \mathbb{R})$ then $v \in \mathcal{C}^{1,2}([0,T] \times \mathbb{R})$ because of (137). That is:

$$V_t = p(t, S_t)\, S_t + q(t, S_t)\, e^{r\,t}\,, \tag{144}$$

and it follows that $V_t = v(t, S_t)$ for $v(t, s) = p(t, s)\, s + q(t, s)\, e^{r\,t}$ and $v \in \mathcal{C}^{1,2}([0,T] \times \mathbb{R})$.

and by Itô's lemma we have that

$$dV_t = \partial_s v(t,S_t)\,dS_t + \left(\partial_t v(t,S_t) + \frac{1}{2}\partial_{ss} v(t,S_t)\sigma^2\,S_t^2\right)dt\,, \tag{157}$$

and given that v satisfies the PDE in (150) then,

$$\partial_t v(t,S_t) + \frac{1}{2}\partial_{ss} v(t,S_t)\sigma^2\,S_t^2 = v(t,S_t)\,r - \partial_s v(t,S_t)\,S_t\,r\,, \tag{158}$$

and thus, (157) becomes

$$\begin{aligned} dV_t &= \partial_s v(t,S_t)\,dS_t + (v(t,S_t)\,r - \partial_s v(t,S_t)\,S_t\,r)\,dt && (159)\\ &= \partial_s v(t,S_t)\,dS_t + (v(t,S_t) - S_t\,\partial_s v(t,S_t))\,e^{-r\,t}\,e^{r\,t}\,r\,dt && (160)\\ &= \phi_t\,dS_t + \psi_t\,B_t\,r\,dt && (161)\\ &= \phi_t\,dS_t + \psi_t\,dB_t\,, && (162) \end{aligned}$$

and proves that (ϕ_t, ψ_t) is a self-financed trading strategy. Then, given that v has terminal condition given by $v(T,S) = f(S)$ we have that

$$V_T = v(T,S_T) = f(S_T)\,, \tag{163}$$

and concludes the proof. ■

5.4 The Black-Scholes formula for a European Call option

Theorem 5.5 *For a standard European call option $f(S) = \max(S-K,0) = (S-K)^+$ and in this case, the Black-Scholes PDE has an explicit solution given by the famous **Black-Scholes formula**:*

$$C(t,S) = C^{BS}(t,S,K,\sigma,r) = S\,\Phi(d_1) - K\,e^{-r\,(T-t)}\Phi(d_2)$$

where $\tau = T-t$ is the time-to-maturity and

$$d_1 \quad = \quad \frac{\log\frac{S}{K} + (r+\frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}} \quad , \quad d_2 \quad = \quad d_1 - \sigma\sqrt{\tau}$$

where $\Phi(x) = \int_{-\infty}^{x} \frac{e^{-z^2/2}}{\sqrt{2\pi}}dz$ is the standard cumulative Normal distribution function. Note that the call price C^{BS} actually depends on five parameters, but only S and t change over time.

We have shown how to replicate a terminal payoff of $f(S_T) = \max(S_T - K, 0)$ under the Black-Scholes model. The steps are:

(i) With $V_0 = C(0, S_0)$ dollars in cash at time zero we can set up the self-financed portfolio at time 0 with $\partial_S C(0, S_0)$ units of the underlying and $C(0, S_0) - \partial_S C(0, S_0)\, S_0$ units of the bond.

(ii) Dynamically trading the share from 0 to T, holding $\phi_t = \partial_S C(t, S_t)$ units of the share at each time instant t. The quantity $\partial_S C(t, S_t) = C_S(t, S_t)$ is known as the **Delta** (Δ) of the option at time t.

(iii) Hold our remaining wealth $V_t - C_S(t, S_t)\, S_t$ at each time instant t in the risk-free bank account.

(iv) Then, at maturity, our self-financed portfolio has value $V_T = f(S_T)$.

Thus $V_0 = C(0, S_0)$ is the cost of replicating $f(S_T)$, since everything we do after time zero is self-financed. If the option price in the market $C^{mkt} > V_0 = C(0, S_0)$, then there is **arbitrage**. Specifically, we can sell the option in the market for C^{mkt} and replicate it using the arguments above at a cost of $C(0, S_0)$ to realise a risk-less profit of $C^{mkt} - C(0, S_0) > 0$. Conversely, if the option is too cheap in the market, that is $C^{mkt} < V_0 = C(0, S_0)$, we can buy the option, and replicate $-f(S_T)$ at a cost of $-C(0, S)$, and again realise a risk-less profit of $C(0, S_0) - C^{mkt} > 0$. Thus $C(0, S_0)$ is the **unique no-arbitrage price** of the option at time 0. Similarly, $C(t, S_t)$ is the **unique no-arbitrage price** of the option at time t, if we re-do the argument above but start at time t rather than time zero. The arguments above hold for any payoff $f(S_T)$.

IMPORTANT: Note that the Black-Scholes PDE and boundary condition are **independent of** μ, and hence the no-arbitrage price of the option is also independent of μ, as for the binomial model.

Thus $C(t, S_t) = C^{BS}(t, S_t, K, \sigma, r)$ and $\phi_t = C_S(t, S_t)$, which can be calculated explicitly using the formula in (164) below. We can also compute $C(t, S)$ explicitly if the terminal payoff is $\log S_T$, $(\log S_T)^2$ or S_T^p (see mock exams and homeworks), in which case the argument is exactly the same except the boundary condition for the PDE will change and thus $C(t, S)$ will change, and hence so will $\phi_t = C_S(t, S_t)$.

5.4.1 Numerical example

Assume current share price is 1, volatility is .10 and interest rate is .05. Price a call option at time zero with strike 1.1 with maturity 1: take $S = 1, K = 1.1, \sigma = .1, \tau = 1, r = .05$ and $t = 0$ and $\tau = T$. Plugging these numbers into the BS formula we obtain

$$\begin{aligned} d_1 &= -0.40310179804324886\,, \\ d_2 &= -0.5031017980432488\,, \end{aligned}$$

and the call price

$$C = 0.02173945155462853\,.$$

See the GitHub repository of the course for the implementation of this formula in Python.

We can easily verify that $C(t,S) \to \max(S-K,0)$ as $t \to T$, and (less obvious) $C(t,S) \to S$ if $T \to \infty$. It can be shown from something called **Jensen's inequality** that $C(t,S)$ is increasing in the time-to-maturity $\tau = T - t$.

Remark 5.1 *The initial (i.e. $t=0$) cost of the replicating strategy at time zero is $C(0,S_0)$, and at each time instant, $C(t,S_t)$ is the unique no-arbitrage price of the call option (see next subsection to see why this is so).*

5.5 The Greeks

The quantity $C_{SS}(t,S_t)$ is known as the **Gamma** (Γ) and $C_t(t,S_t)$ is known as the **Theta** (Θ) of the option, and as we shall see, these partial derivatives have explicit formulae for the case when $f(S) = \max(S-K,0) = (S-K)^+$, i.e. a European call option.

The partial derivatives of the BS formula with respect to some of the parameters are

$$\begin{aligned} \Delta &= \frac{\partial C}{\partial S} = \Phi(d_1) > 0, \\ \Gamma &= \frac{\partial^2 C}{\partial S^2} = \frac{\partial \Delta}{\partial S} = \frac{n(d_1)}{S\sigma\sqrt{\tau}} > 0\,, \\ \nu &= \frac{\partial C}{\partial \sigma} = Sn(d_1)\sqrt{\tau} > 0 \end{aligned} \tag{164}$$

where $n(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ is the standard Normal density. Δ, Γ and ν are known as the Delta, Gamma and Vega respectively of the option. The proof of these expressions for the Greeks are very tedious and not examinable.

(i) Delta measures the responsiveness of the call option price to small changes in the underlying share price.

(ii) Vega measures the responsiveness of the call option price to small changes in the volatility.

(iii) Gamma measures the responsiveness of the Delta to small changes in the underlying share price.

For the numerical example above, we obtain $\Delta = 0.343436668962647$, $\Gamma = 3.678117344101192$ and $\nu = 0.3678117344101192$.

5.6 The Feynman-Kac formula

Theorem 5.6 *Let $v(t,S)$ be a solution to the following PDE*

$$v(t,S)\,r - \partial_t v(t,S) - \partial_s v(t,S)\,S\,r - \frac{1}{2}\partial_{ss} v(t,S)\sigma^2\,S^2 = 0\,, \tag{165}$$

with terminal condition $v(T,S) = f(S)$. Then, $v(t,S)$ admits the following probabilistic representation

$$v(t,S) = e^{-r\,(T-t)}\,\mathbb{E}^{\mathbb{Q}}\left[f(S_T)\middle|S_t = S\right]\,, \tag{166}$$

for all $(t, S) \in [0, T] \times \mathbb{R}$ *and where*

$$dS_t = r\, S_t\, dt + \sigma\, S_t\, dW_t^{\mathbb{Q}}, \qquad S_0 > 0\,. \tag{167}$$

It follows from Theorem 5.6 that $C(t, S)$ (the price of a call option which pays $(S_T - K)^+$ at time T under the Black-Scholes model) also has the following *probabilistic* representation:

$$C(t, S) \quad = \quad e^{-r\,(T-t)} \mathbb{E}^{\mathbb{Q}} \left[(S_T - K)^+ \,|\, S_t = S \right]$$

where $\mathbb{Q}$ is a new probability measure under which S satisfies

$$dS_t = S_t \left(r\, dt + \sigma\, dW_t^{\mathbb{Q}} \right), \tag{168}$$

and $W^{\mathbb{Q}}$ is a standard Brownian motion under $\mathbb{Q}$.

In general, Theorem 5.6 implies the following: **the option price (i.e. the cost of replicating the option) is the discounted expected value of $f(S_T)$ in the risk-neutral world $\mathbb{Q}$ where the drift is r not μ**. We refer to this world as the **risk neutral measure**. Note that (168) looks similar to the real-world Black-Scholes SDE $dS_t = \mu\, S_t\, dt + \sigma\, S_t\, dW_t$, but now written in a world where $dS_t = r\, S_t\, dt + \sigma\, S_t\, dW_t^{\mathbb{Q}}$ in which the drift is r and not μ.

Example 5.2 *Price a digital call option under the Black-Scholes model which pays 1 if $S_T > K$ and zero otherwise.*

Solution: By the same hedging argument as for European options, the price $P(t, S)$ of the digital call satisfies also satisfies the Black-Scholes PDE but with boundary condition $f(S) = 1_{S>K}$. Then from the Feynman-Kac formula, $P(t, S)$ has the probabilistic representation

$$\begin{aligned} P(S, t) \quad &= \quad e^{-r(T-t)} \mathbb{E}^{\mathbb{Q}} \left[1_{S_T > K} \,|\, S_t = S \right] \\ &= \quad e^{-r(T-t)} \mathbb{Q}(S_T > K \,|\, S_t = S) \end{aligned}$$

where (again) S_t satisfies the SDE $dS_t = S_t(rdt + \sigma dW_t^{\mathbb{Q}})$ under the probability measure $\mathbb{Q}$, and $\mathbb{Q}(A)$ denotes the probability of an event A under the probability measure $\mathbb{Q}$, and we have used that for any random variable X, $\mathbb{P}(X > K) = \mathbb{E}(1_{X>K})$. If we now let $t = 0$, then we can compute $\mathbb{Q}(S_T > K)$ similar to before as:

$$\begin{aligned} \mathbb{Q}(S_T > K) \quad &= \quad \mathbb{Q}\left(\log \frac{S_T}{S_0} > \log \frac{K}{S_0} \right) \\ &= \quad \mathbb{Q}\left(\frac{\log \frac{S_T}{S_0} - (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} > \frac{\log \frac{K}{S_0} - (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} \right) \\ &= \quad \mathbb{Q}\left(Z > \frac{\log \frac{K}{S_0} - (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} \right) \\ &= \quad \Phi^c(z) \end{aligned}$$

where

$$z = \frac{\log \frac{K}{S_0} - (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}},$$

note that we have now just replaced μ with r. Similarly

$$\mathbb{Q}(S_T > K \mid S_t = S) \quad = \Phi^c(z)$$

where now

$$z = \frac{\log \frac{K}{S_t} - (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}\,.$$

Note that

$$\begin{aligned} S_T &= S_0 e^{(r-\frac{1}{2}\sigma^2)T + \sigma W_T^{\mathbb{Q}}} && (169)\\ &= S_0 e^{(r-\frac{1}{2}\sigma^2)t + \sigma W_t} e^{(r-\frac{1}{2}\sigma^2)(T-t) + \sigma(W_T^{\mathbb{Q}} - W_t^{\mathbb{Q}})} && (170)\\ &= S_t e^{(r-\frac{1}{2}\sigma^2)(T-t) + \sigma(W_T^{\mathbb{Q}} - W_t^{\mathbb{Q}})} && (171) \end{aligned}$$

so at time t we have $\log S_T \sim N(\log S_t + (r - \frac{1}{2}\sigma^2)(T-t), \sigma^2(T-t))$.

Example 5.3 *Price a long position on a forward with strike K and maturity T; here, the holder of the contract will buy the asset at time T for price K. The payoff at time T of such derivative is given by*

$$S_T - K\,. \tag{172}$$

Solution: By Feynman-Kac the price is given by

$$\begin{aligned} P(S,t) &= e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\left[S_T - K \mid S_t = S\right] \\ &= e^{-r(T-t)}\left(S\, e^{r\,(T-t)} - K\right) \\ &= S - K\, e^{-r(T-t)}\,, && (173) \end{aligned}$$

which is a well-known formula that can be derived from first principles – see Chapter 5 in Hull (2003). The above formula at time zero would read as

$$S - K\, e^{-r\,T}\,, \tag{174}$$

and the forward price (which is defined as the strike K that makes the value of the contract equal to zero) is given by

$$S\, e^{r\,T}\,. \tag{175}$$

5.6.1 Implied volatility

The Vega $\frac{\partial C}{\partial \sigma} = Sn(d_1)\sqrt{\tau}$ of a call option under Black-Scholes is positive, so C is monotonically increasing as a function of σ. Thus, given an observed call price C^{obs} in the market, we can extract a unique σ value consistent i.e. such that

$$C(S, K, \hat{\sigma}, \tau, r) \quad = \quad C^{obs}\,.$$

if $\max(S_0 - Ke^{-rT}, 0) \leq C^{obs} < S_0$. This σ is known as the **implied volatility** of the option, and is a very important concept in practice.

5.7 Continuous-time martingales

Martingales are stochastic processes that appear often in mathematical finance. They are used to model 'fair games'. Within pricing and hedging of derivatives, it turns out that there are well-known links between martingales and continuous hedging of contingent claims – see Harrison and Pliska (1981). Next, we provide a short introduction to the topic so that the curious reader has the tools to consult more advanced books in the area. We will see that for the last two chapters we have been working with processes that are martingales.

A continuous-time martingale $(X_t)_{t\geq 0}$ is a stochastic process which satisfies the following two conditions:

(i) $\mathbb{E}[X_t \,|\, \mathcal{F}_s] = X_s$ for $0 \leq s \leq t$, where $\mathcal{F}_s$ is the information available at time s. For our purposes on this course, $\mathcal{F}_s$ just means all the information in the historical sample path of the process X from time 0 to time s.

(ii) $\mathbb{E}[|X_t|] < \infty$ for all finite $t \geq 0$.

We now describe some well known examples of continuous-time martingales:

Example 5.4 ***Brownian motion****. For standard Brownian motion, we know that*

$$W_t - W_s \sim N(0, t-s)\,. \tag{176}$$

Thus

$$\begin{aligned}
\mathbb{E}[W_t \,|\, \mathcal{F}_s] &= \mathbb{E}[W_t - W_s + W_s \,|\, \mathcal{F}_s] \\
&= W_s + \mathbb{E}[W_t - W_s \,|\, \mathcal{F}_s] \\
&= W_s + \mathbb{E}[W_t - W_s] \\
&= W_s\,.
\end{aligned}$$

Moreover, $\mathbb{E}[|W_t|] = \int_{-\infty}^{\infty} |x| \frac{e^{-\frac{x^2}{2t}}}{\sqrt{2\pi t}} dx < \infty$ *because for* $|x|$ *large, the decaying exponential term decays much faster than the* $|x|$ *term is growing. Thus* W_t *is a martingale.*

Example 5.5 *Let* $M_t = W_t^2 - t$*. Then*

$$\begin{aligned}
\mathbb{E}[M_t - M_s \,|\, \mathcal{F}_s] &= \mathbb{E}[W_t^2 - t - (W_s^2 - s) \,|\, \mathcal{F}_s] \\
&= \mathbb{E}[W_t^2 - W_s^2 \,|\, \mathcal{F}_s] - (t-s) \\
&= \mathbb{E}[(W_s + W_t - W_s)^2 - W_s^2 \,|\, \mathcal{F}_s] - (t-s) \\
&= \mathbb{E}[W_s^2 + 2W_s(W_t - W_s) + (W_t - W_s)^2 - W_s^2 \,|\, \mathcal{F}_s] - (t-s) \\
&= \mathbb{E}[2W_s(W_t - W_s) + (W_t - W_s)^2 \,|\, \mathcal{F}_s] - (t-s) \\
&= 2W_s\mathbb{E}[W_t - W_s \,|\, \mathcal{F}_s] + \mathbb{E}[(W_t - W_s)^2 \,|\, \mathcal{F}_s] - (t-s) \\
&\quad \textit{(using that } W_s \textit{ is known at time } s\textit{)} \\
&= 2W_s\mathbb{E}[W_t - W_s] + \mathbb{E}[(W_t - W_s)^2] - (t-s) \\
&= 0 + t - s - (t-s) \\
&\quad \textit{(using that } W_t - W_s \sim N(0, t-s)\textit{)} \\
&= 0
\end{aligned}$$

and hence $\mathbb{E}[M_t|\mathcal{F}_s] = \mathbb{E}[M_s|\mathcal{F}_s] = M_s$. *Moreover,* $\mathbb{E}[|M_t|] \leq \mathbb{E}[W_t^2 + t] = 2t < \infty$, *so* $M_t = W_t^2 - t$ *also satisfies the second condition, and hence* M_t *is a martingale.*

Example 5.6 ***The Black-Scholes share price process with*** $\mu = 0$. *Let* $S_t = S_0 e^{\sigma W_t - \frac{1}{2}\sigma^2 t}$. *Then*

$$\mathbb{E}[S_t \,|\, \mathcal{F}_s] \quad = \quad \mathbb{E}\left[S_0 e^{\sigma W_s - \frac{1}{2}\sigma^2 s} e^{\sigma(W_t - W_s) - \frac{1}{2}\sigma^2(t-s)} \,|\, \mathcal{F}_s\right] \quad = \quad S_s \mathbb{E}\left[e^{\sigma(W_t - W_s) - \frac{1}{2}\sigma^2(t-s)} \,|\, \mathcal{F}_s\right]$$

where we have used that $S_s = S_0 e^{\sigma W_s - \frac{1}{2}\sigma^2 s}$, *and the value of* S_s *is known at time* s. *But for any normal random variable* $X \sim N(\mu, \nu^2)$, *the moment generating function of* X *is given by* $\mathbb{E}(e^{pX}) = e^{\mu p + \frac{1}{2}\nu^2 p^2}$. *In our case here,* $X = \sigma(W_t - W_s) - \frac{1}{2}\sigma^2(t-s)$ *so* $\mu = -\frac{1}{2}\sigma^2(t-s)$ *and* $\nu^2 = \sigma^2(t-s)$ *and* $p = 1$, *so we see that*

$$\begin{aligned}
\mathbb{E}[S_t|\mathcal{F}_s] \quad = \quad \mathbb{E}\left[S_s e^{\sigma(W_t - W_s) - \frac{1}{2}\sigma^2(t-s)} \,|\, \mathcal{F}_s\right] \quad &= \quad S_s e^{\mu p + \frac{1}{2}\nu^2 p^2} \\
&= \quad S_s e^{-\frac{1}{2}\sigma^2(t-s) + \frac{1}{2}\sigma^2(t-s)} \\
&= \quad S_s \,.
\end{aligned}$$

Moreover, setting $s = 0$ *we see that* $\mathbb{E}[|S_t|] = \mathbb{E}[S_t] = S_0 < \infty$, *so* S_t *is a martingale.*

Corollary 5.7 *Recall that under* $\mathbb{Q}$ *the asset price follows*

$$dS_t = r\, S_t\, dt + \sigma\, S_t\, dW_t^{\mathbb{Q}}\,, \tag{177}$$

it then follows that

$$Y_t = e^{-r\,t}\, S_t\,, \tag{178}$$

is a martingale under $\mathbb{Q}$. *That is, the discounted asset price is a martingale under* $\mathbb{Q}$.

6 Pricing with simulations

In this section we assume we are able to generate independent and identically distributed (i.i.d.) samples of a uniform $U([0,1])$.

Proposition 6.1 *Let X be a random variable with distribution function $F_X : \mathbb{R} \to [0,1]$, then*

$$F_X^{-1}(U) \sim X\,. \tag{179}$$

We recall Proposition 2.11 that said that if one has a collection $(X_i)_{i\in\mathbb{N}}$ of i.i.d. samples from a random variable X with mean μ satisfying some integrability conditions, then the sample average

$$\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i\,, \tag{180}$$

tends to the true expectation μ of X_i as $n \to \infty$. This result is very important for what follows.

6.1 Computing integrals using simulations

Definition 6.1 *Let $f : [a,b] \to R$. Let $\{x_i\}_{i\in\{1,\ldots,N\}}$ be N random numbers uniformly distributed in the interval $[a,b]$. The Monte Carlo estimate for the integral*

$$\int_a^b f(x)dx \tag{181}$$

is given by

$$\frac{b-a}{N}\sum_{i=1}^{N} f(x_i)\,. \tag{182}$$

Example 6.1 *Suppose we wanted to compute the integral*

$$\int_1^5 x^3\,dx\,. \tag{183}$$

From first year calculus we know the above should give

$$\frac{5^4 - 1^4}{4} = 156\,, \tag{184}$$

and using the algorithm above yields 156.02. Below, an extract of the code to produce the result.

```python
nsims = 10_000_000
U = np.random.rand(nsims,1) # Uniform values in (0,1)
X = 4*U + 1 # Uniform values in (1,5)
exact = (5**4 - 1**4)/4 # exact value of the integral
simulation = (5-1)/nsims * np.sum(X**3)  # approximation
```

So, what about infinite integrals? We can perform a change of variables and transform the infinite integral to an integral over an open interval.

Definition 6.2 *Let X be a random variable with distribution function $F_X : \mathbb{R} \to [0,1]$, then one can approximate its expectation as follows: let $\{U_i\}_{i\in\{1,\dots,N\}}$ be N independent and identically distributed uniform random variables, then by letting $X_i = F_X^{-1}(U_i)$ we have that $\{X_i\}_{i\in\{1,\dots,N\}}$ are independent and identically distributed copies of X. The Monte Carlo estimate for the expectation of $g(X)$ is*

$$\mathbb{E}[g(X)] \approx \frac{1}{N}\sum_{i=1}^{N} g(X_i)\,. \tag{185}$$

Example 6.2 *Let $X \sim exp(\lambda)$. We know that*

$$F_X(x) = 1 - e^{-\lambda\, x}\,, \qquad x \geq 0\,, \tag{186}$$

and a short calculation gives

$$F_X^{-1}(y) = -\frac{\log(1-y)}{\lambda}\,, \qquad y \in (0,1)\,. \tag{187}$$

By using integration by parts we can explicitly compute

$$\mathbb{E}\left[X^n\right] = \frac{n!}{\lambda^n}\,, \tag{188}$$

and we can use the method above to see how close we get to the actual expectation when employing simulations.

Below, there is code that replicates the steps above for $\lambda = 2.5$ and $n = 2$. We employ 10,000,000 simulations.

```
def F_inv_exp(x, l = 2.5):
    return -np.log(1.-x) / l

nsims = 10_000_000 #number of simulation
l = 2.5 #value of lambda
U = np.random.rand(nsims,1) #generate nsims uniforms
X = F_inv_exp(U, l) #generate nsims exponentials

n = 2
simulation = np.sum(X**n)/nsims #approximation using simulations
exact = factorial(n)/ (l**n) #this is the actual value
```

In a run of the above, the following is obtained: the exact calculation is 0.32 and using simulations gives 0.32012. If we repeat the experiment but now with $\lambda = 0.5$ and $n = 4$, i.e., we compute $\mathbb{E}[X^4]$, we obtain that the exact calculation is 384.0 and using simulations gives 383.51154.

Can we use the techniques above to price derivatives?

6.2 Pricing derivatives

So far we have learnt that the risk neutral price of a derivative is its discounted expected value under a given risk neutral measure $\mathbb{Q}$ – see results from the Feynman-Kac section. To price a European-style derivative with payoff function $f : \mathbb{R}^+ \to \mathbb{R}$ the Monte Carlo unbiased estimate for the price of the option at time t would be

$$e^{-r(T-t)}\,\frac{1}{N}\sum_{i=1}^{N} f(S_T^{(i)}) \tag{189}$$

where $\{S_T^{(i)}\}_{i\in\{1,\dots,N\}}$ are independent samples of S_T under $\mathbb{Q}$.

Let us start by computing the Black-Scholes price for a European call option. From the previous sections we know that the formula is given by

$$C^{BS}(S,K,\sigma,T-t,r) = S\Phi(d_1) - Ke^{-r\tau}\Phi(d_2)$$

where $\tau = T-t$ is the time-to-maturity and

$$d_1 \;=\; \frac{\log\frac{S}{K} + (r+\frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}} \quad , \quad d_2 \;=\; d_1 - \sigma\sqrt{\tau}$$

where $\Phi(x) = \int_{-\infty}^{x} \frac{e^{-z^2/2}}{\sqrt{2\pi}}dz$ is the standard cumulative Normal distribution function. In Python, this function would look like:

```
def computeBlackScholesCallPrice(t,T,S,r,sigma,K):
    d1 = (np.log(S/K) + (r + 0.5*sigma**2)*(T-t))/(sigma * np.sqrt(T-t))
    d2 = d1 - sigma* np.sqrt(T-t)
    return S*stat.norm.cdf(d1) - K*np.exp(-r*(T-t))*stat.norm.cdf(d2)
```

Let $t \in [0,T]$, and let $S_t = S$, then

$$S_T = S_t\, e^{\left(r-\frac{1}{2}\sigma^2\right)(T-t)+\sigma\,(W_T-W_t)}\,, \tag{190}$$

under $\mathbb{Q}$, where W is a Brownian motion under $\mathbb{Q}$. Thus,

$$S_T \sim S\, e^{\left(r-\frac{1}{2}\sigma^2\right)(T-t)+\sigma\,\sqrt{T-t}\,Z}\,, \tag{191}$$

where Z is a standard normally distributed random variable. Thus, the following algorithm produces simulations of S_T under $\mathbb{Q}$ given that $S_t = S$:

```
def simulate_S_T_under_Q(t,T,S,r,sigma,nsims):
    Z = np.random.randn(nsims,1)
    exponent = (r-0.5*sigma**2)*(T-t) + sigma*np.sqrt(T-t)*Z
    return S*np.exp(exponent)
```

Then, an algorithm to obtain the Black-Scholes price for a European call option would be:

```python
# Model parameters
t = 0
T = 1
S = 100
r = 0.03
sigma = 0.1
K = 100
nsims = 10_000_000

S_T = simulate_S_T_under_Q(t,T,S,r,sigma,nsims)

def call_option_payoff(K, S):
    return np.maximum(S-K,0)

approximation = np.exp(-r*(T-t))*np.sum(call_option_payoff(K, S_T))/nsims
exact = computeBlackScholesCallPrice(t,T,S,r,sigma,K)
```

Running the above yields that the exact price is 5.5818 and the approximation is 5.5815.

6.3 Pricing a path dependent derivative

We can also use the techniques described above to price more complex derivatives.

Definition 6.3 *(Knock-out barrier options) A knock-out barrier option becomes worthless if the underlying share price hits a barrier. If the underlying share price begins below the barrier, the option is referred as up-and-out. Similarly, if the asset price starts above the barrier, then it is called down-and-out. The payoff in case it does not knock-out is typically that of a call or a put option. The four combinations we will study are:*

(i) down-and-out call option,

(ii) down-and-out put option,

(iii) up-and-out call option,

(iv) up-and-out put option.

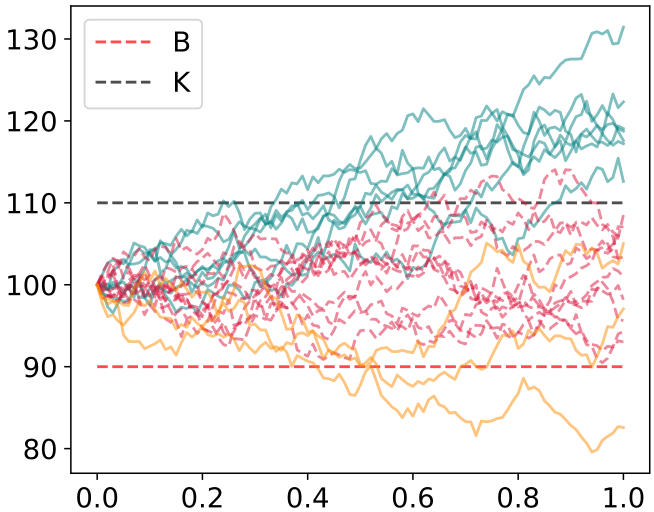

Figure 15: Here we plot 20 trajectories of S_t under $\mathbb{Q}$ and we colour the paths as follows, (i) teal if S never went below B and ended above K at time T, (ii) red if S never went below B but it did not end above K at time T, and (iii) orange if S went below B. Parameters are $S_0 = 100$, $r = 0.03$, $\sigma = 0.1$, $K = 110$, $B = 90$, and $T = 1$.

Proposition 6.2 *To price a knock-out barrier options using simulations we:*

(i) simulate trajectories for $(S_t)_{t\in[0,T]}$ under $\mathbb{Q}$,

(ii) compute the payoff of each path – zero if in the given path the option knocked-out, and the payoff at T (that of the call or the put) if the option did not knock-out,

(iii) average payoffs across simulations and discount.

The following code produces the pricing of a down-and-out call option with $S_0 = 100$, $\mu = 0.01$, $r = 0.03$, $\sigma = 0.1$, $K = 100$, $B = 90$, and $T = 1$.

```python
# Model parameters
t = 0
T = 1
S = 100
r = 0.03
sigma = 0.1
K = 100
B = 90
nsims = 100_000

GBM = utils.GeometricBrownianMotion(x0 = S, mu = r, sigma = sigma, T = T, Nt = 100)
paths = GBM.simulate(nsims = nsims)
payoff = call_option_payoff(K, paths[-1,:]) * (np.min(paths,axis = 0) > B)
```

```
approximation = np.exp(-r*(T-t))*np.sum(payoff)/nsims
```

In the code above we used a class called `GeometricBrownianMotion`; this class is coded as follows:

```
import numpy as np

class GeometricBrownianMotion:
    """
    Model parameters for the environment.
    """
    def __init__(self, x0 = 100., mu = 0.05, sigma = 0.1, T = 1.0, Nt = 100):
        self.x0 = x0
        self.mu = mu
        self.sigma = sigma
        self.T = T
        self.Nt = Nt
        self.timesteps = np.linspace(0, self.T, num = (Nt+1))

    def simulate(self, nsims=1):
        x = np.zeros((self.Nt+1, nsims))
        x[0,:] = self.x0
        dt = self.T/(self.Nt)
        errs = np.random.randn(self.Nt, nsims)
        for t in range(self.Nt):
            x[t + 1,:] = x[t,:] * np.exp( dt * (self.mu - 0.5*self.sigma**2) +
            ↪  np.sqrt(dt) * self.sigma * errs[t,:] )
        return x
```

The above calculation produces a price of 5.53, which is smaller than that of the previous example. This makes sense since the payoff of the call option above dominates that of the down-and-out call option with the same strike.

7 Barrier options (not examinable)

In this section we study barrier options. Two of the most common types of barrier options are: know-out and knock-in. Knock-out barrier option are worthless if the asset price touches the barrier and knock-in barrier options are worthless unless the asset price touches the barrier. Depending on whether the barrier is above or below the current level of the asset, we coin the barrier options as: (i) up-and-out, (ii) down-and-out, (iii) up-and-in, and (iv) down-and-in. Here, (i) and (ii) are knock-out options where in (i) the barrier is above the current asset price and in (ii) the barrier is below. Similarly, (iii) and (iv) are knock-in barrier options, in (iii) the barrier is above the current asset price and in (iv) the barrier is below.

7.1 Semi-static hedging: pricing and hedging barrier options

We begin with an equation that will be key to develop the semi-static hedging arguments.

Proposition 7.1 *Let $\mathbb{Q}$ be the risk neutral measure. Under the Black-Scholes model, for the payoff function f, the following holds true*

$$e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\left[f(S_T)\,\Big|\,\mathcal{F}_t\right] \quad = \quad e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\left[\left(\frac{S_T}{S_t}\right)^{2\gamma} f\left(\frac{S_t^2}{S_T}\right)\Big|\mathcal{F}_t\right] \tag{192}$$

where

$$\gamma = \frac{1}{2} - r/\sigma^2\,. \tag{193}$$

Proof. This proof is out of the scope of the course because it makes use of something called Girsanov theorem. I provide the proof in Appendix 8.3 for the curious reader. ■

Using this relation we can replicate and hence price a knock-out call option with strike K and barrier $B < K$ with $B < S_0$ with European type contracts.

Remark 7.1 *Let $f(S) = \max(S - K, 0)$ and assume that $S_t = B$ for some time $t \in [0, T]$. Then, (192) says that an European call option at time t with strike K is worth the same as an option that pays*

$$\left(\frac{S_T}{B}\right)^{2\,\gamma} \max\left(\frac{B^2}{S_T} - K, 0\right)\,, \tag{194}$$

at time T.

Proposition 7.2 *The price of a knock-out call option with strike K and barrier $B < K$ with $B < S_0$ is*

$$e^{-r\,T}\left\{\mathbb{E}^{\mathbb{Q}}\left[(S_T - K)^+\right] - \mathbb{E}^{\mathbb{Q}}\left[\left(\frac{S_T}{B}\right)^{2\,\gamma} \max\left(\frac{B^2}{S_T} - K, 0\right)\right]\right\}\,. \tag{195}$$

Proof. To replicate a knock-out call option with strike K and barrier $B < K$ with $B < S_0$ one follows:

(i) At time $t = 0$: buy a call option with strike K.

(ii) At time $t = 0$: sell an option with payoff function

$$\left(\frac{S_T}{B}\right)^{2\gamma} \max\left(\frac{B^2}{S_T} - K, 0\right). \tag{196}$$

Note that generally this is not a traded option, but we can sell *the replicating portfolio*.

(iii) If S_t ever hits B: then the value of (i) plus the value of (ii) is zero as a consequence of Proposition 7.1; thus, we can unwind the position at no cost. Alternatively, if S_t never hits B, then the value of (i) plus the value of (ii) at T is exactly that of (i) because (ii) is worth zero. This is because $B < S_T$ and thus $B^2/S_T < B < K$ which means that

$$\max\left(\frac{B^2}{S_T} - K, 0\right) = 0, \tag{197}$$

and shows that in this case (ii) is worth zero.

Following the steps above replicates the payoff of the down-and-out knock out barrier call option. Thus, its price is (195). ∎

The above is known as **semi-static hedging**.
Note how by using European-type contracts we are able to mimic the behaviour of the payoff of the down-and-out knock out barrier call option.

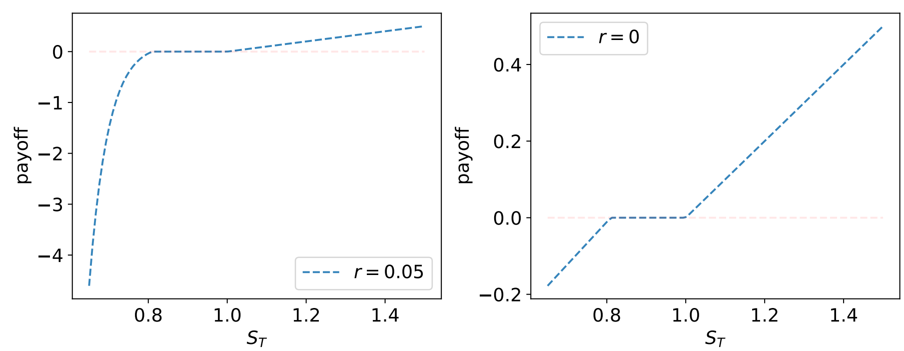

Figure 16: Here we have plotted the semi-static hedge portfolio payoff $(S-K)^+ - (\frac{S}{B})^{2\gamma}\max(\frac{B^2}{S} - K, 0)$ as a function of S for a down-and-out call with $K = 1$, $B = 0.9$, and $r = 0.05$ (left) and $r = 0$ (right).

Corollary 7.3 *The price of a down-and-out knock out barrier call option is given by*

$$\begin{aligned} P(S_0, 0) &= e^{-rT}\mathbb{E}^{\mathbb{Q}}\left[(S_T - K)^+ - \left(\frac{S_T}{B}\right)^{2\gamma} \max\left(\frac{B^2}{S_T} - K, 0\right)\right] && (198) \\ &= e^{-rT}\int_0^\infty \left(f(S) - \left(\frac{S}{B}\right)^{2\gamma} f\left(\frac{B^2}{S}\right)\right) p_{S_T}(S)\, dS && (199) \end{aligned}$$

where $p_{S_T}(S)$ is the (lognormal) density of S_T under $\mathbb{Q}$, and the price is less than the price of the European as expected. If $r = 0$, then $2\gamma = 1$, and the payoff of the contract we sell simplifies to

$$\frac{S_T}{B}\max\left(\frac{B^2}{S_T} - K, 0\right) \quad = \quad \max\left(B - \frac{KS_T}{B}, 0\right) \quad = \quad \frac{K}{B}\max\left(\frac{B^2}{K} - S_T, 0\right)$$

i.e. $\frac{K}{B}$ *European put options with strike* $\frac{B^2}{K}$.

Remark 7.2 *In the example above $B < K$. If $K < B$ the semi-static hedging argument still works as long as we set $f(S) = (S - K)^+ 1_{S>B}$.*

Proposition 7.4 *For a No-Touch option, which pays 1 at time T if S stays above B over $[0, T]$ the semi-static hedging can be done using*

$$f(S) = 1_{S>B}\,. \tag{200}$$

Note that

$$f\left(\frac{B^2}{S}\right) = 1_{S<B}\,. \tag{201}$$

Hence we can semi-statically hedge the No-Touch option by buying a digital call option with strike B and selling a contract which pays $(S_T/B)^{2\gamma}\, 1_{S_T<B}$.

Remark 7.3 *If we want to compute the vega of the No-Touch at time zero, we replace $p_{S_T}(S)$ with $\frac{\partial}{\partial\sigma}p_{S_T}(S)$ in (199). Similarly if we want the delta at time zero, we replace $p_{S_T}(S)$ with $\frac{\partial}{\partial S_0}p_{S_T}(S)$.*

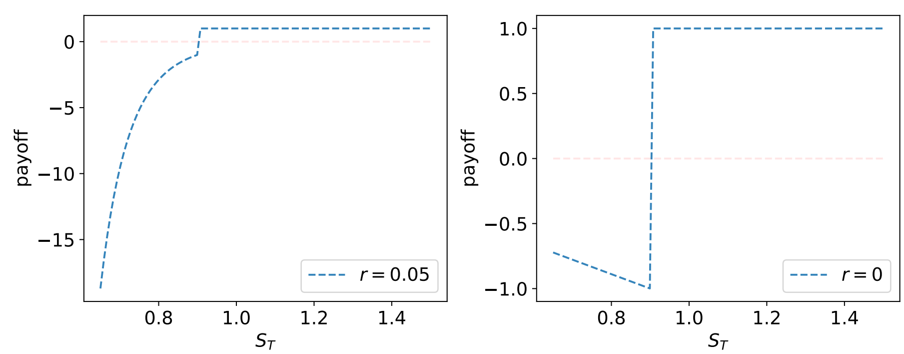

Figure 17: Here we have plotted the semi-static hedge portfolio payoff $1_{S>B} - (\frac{S}{B})^{2\gamma}\, 1_{S<B}$ as a function of S for the No-Touch option above with parameters $B = 0.9$, and $r = 0.05$ (left) and $r = 0$ (right).

7.2 PDE approach: pricing and hedging barrier options

Consider a *time-window barrier option* that pays $\max(S_T - K, 0)$ at time T if S_t stays above $B < K$ (with $B < S_0$) over the interval $[0, T_2]$ where $0 < T_2 < T$, and knock out (i.e. dies otherwise). Let $P(S, t)$ be the price of this time-window barrier option at time $t \le T$, with the convention that $P(S, t)$ is the price when the derivative has not been triggered before.

From the Black-Scholes hedging/replication arguments above, we know that the no-arbitrage price $P(S, t)$ satisfies the Black-Scholes PDE:

$$P_t(S,t) + rP_S(S,t)S + \frac{1}{2}P_{SS}(S,t)S^2\sigma^2 \quad = \quad rP(S,t)$$

with terminal condition $P(S, T) = (S - K)^+$, but we now have the additional boundary condition $P(B, t) = 0$ for $t \in [0, T_2]$.

To replicate the contract, we replicate the payoff $(S_T - K)^+ 1_{\underline{S}_{T_2} > B}$ at time T (where $\underline{S}_t := \min_{0 \le u \le t} S_u$) as follows:

(i) Let $(V_t)_{t \ge 0}$ denote total wealth process of the hedging strategy at time t and we require initial wealth $V_0 = v(0, S_0)$.

(ii) Hold $\phi_t = V_S(t, S_t)$ units of the share until time $T \wedge H_B$ where H_B is the hitting time of S to B, and place remaining wealth $V_t - \phi_t S_t$ in the bonds account.

(iii) V then evolves as

$$dV_t \quad = \quad \phi_t \, dS_t + (V_t - \phi_t \, S_t) \, r \, dt$$

whose solution is then given by $V_t = v(t, S_t)$ (proof not required, comes from the usual Black-Scholes hedging argument), so in particular $V_T = (S_T - K)^+ 1_{\underline{S}_{T_2} > B}$, which hedges a long position in the original contract.

Then, $V_t = v(t, S_t)$; we see in particular that if S hits the barrier at time $\tau_B \in [0, T_2]$, then at this exact moment $V_{\tau_B} = v(\tau_B, B) = 0$ (from the boundary condition we have imposed for the PDE), so we have also replicated the Time Window Barrier option in this scenario.

7.3 Simulations and the advantage of the the semi-static argument

In Section 6 we learnt to price path-dependent derivatives using simulations. Here, I compare the two approaches. For this I will price a down-and-out barrier call option with the following model parameters: $S_0 = 100$, $r = 0.02$, $\sigma = 0.1$, $K = 100$, $B = 90$, and $T = 1$. Below, we recall the approach taken in Section 6.

```
t = 0
T = 1
S = 100
r = 0.03
sigma = 0.1
K = 100
```

```
B = 90
nsims = 1_000_000
GBM = utils.GeometricBrownianMotion(x0 = S, mu = r, sigma = sigma, T = T, Nt = 100)
paths = GBM.simulate(nsims = nsims)
payoff = call_option_payoff(K, paths[-1,:]) * (np.min(paths,axis = 0) > B)
approximation = np.exp(-r*(T-t))*np.sum(payoff)/nsims
```

Next, we employ the techniques described above and we price a derivative that pays

$$e^{-rT}\mathbb{E}^{\mathbb{Q}}\left[(S_T - K)^{+} - \left(\frac{S_T}{B}\right)^{2\gamma} \max\left(\frac{B^2}{S_T} - K, 0\right)\right]. \tag{202}$$

```
def f_func(s):
    return  np.maximum(s-K,0)

def simulate_S_T_under_Q(t,T,S,r,sigma,nsims):
    Z = np.random.randn(nsims,1)
    exponent = (r-0.5*sigma**2)*(T-t) + sigma*np.sqrt(T-t)*Z
    return S*np.exp(exponent)

def gamma_val(r,sigma):
    return 1/2 - r/(sigma**2)

def payoff_two_derivatives(S, K, B, gamma, f: callable):
    return f(S) - (S/B)**(2*gamma)*f(B**2/S)

S_T = simulate_S_T_under_Q(t,T,S,r,sigma,nsims)
gamma = gamma_val(r, sigma)
Payoff = payoff_two_derivatives(S_T, K, B, gamma, f_func)
approximation2 = np.exp(-r*(T-t)) * np.sum(Payoff)/nsims
```

The above two approaches give 5.53 and 5.50 respectively.

References

Cohen, S. N. and Elliott, R. J. (2015). *Stochastic calculus and applications*, volume 2. Springer.

Harrison, J. M. and Pliska, S. R. (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic processes and their applications*, 11(3):215–260.

Hull, J. C. (2003). *Options futures and other derivatives.* Pearson Education India.

Karatzas, I. and Shreve, S. E. (1991). *Brownian motion and stochastic calculus*, volume 113. Springer Science & Business Media.

Protter, P. E. (2005). *Stochastic Integration and Differential Equations*, volume 21. Springer Science & Business Media.

Shreve, S. (2004). *Stochastic calculus for finance II: Continuous-time models.* Springer.

Williams, D. (1991). *Probability with martingales.* Cambridge university press.

8 Appendix

8.1 A note on continuous compounding

Continuous compounding arises as a limit on the number of periods in which interest rates compound within a given time window. We provide an example. Let us fix an initial capital of $c =$£ 100 for illustration purposes and let $R = 5\%$. Consider a contract that pays the following: if the initial investment is not withdrawn in 1 year the contract pays $c\,(1+R)$, alternatively, if the investment is withdrawn, the holder receives c. It is clear that the wealth of the investor is a step function with a jump at time $T = 1$ from £100 to £105. Now, let us study a contract that compounds in the middle of the time interval $[0,1]$ and achieves the same 5% increase at the of $[0,T]$. That is, from $[0,T/2)$ the wealth of the investor is £ 100, then from $[T/2,T)$ the wealth is £ $100\,(1+r^{(2)}/2)$, and at time T it jumps to £ $100\,(1+r^{(2)}/2)^2$. What is the interest rate $r^{(2)}$ such that

$$c\left(1+\frac{r^{(2)}}{2}\right)^2 = c\,(1+R)\,? \tag{203}$$

The answer can be computed explicitly and is given by

$$r^{(2)} = 2\left(\exp\left(\frac{\log(1+R)}{2}\right)-1\right)\,. \tag{204}$$

Similarly, we can ask the same question for $n \in \mathbb{N}$, here we look for $r^{(n)}$ such that

$$\left(1+\frac{r^{(n)}}{n}\right)^n = (1+R)\,, \tag{205}$$

where I have omitted c since it does not play a role in the determination of $r^{(n)}$. Here, the solution is given by

$$\begin{aligned} r^{(n)} &= n\left(\exp\left(\frac{\log(1+R)}{n}\right)-1\right) && (206)\\ &= n\left((1+R)^{1/n}-1\right)\,. && (207)\end{aligned}$$

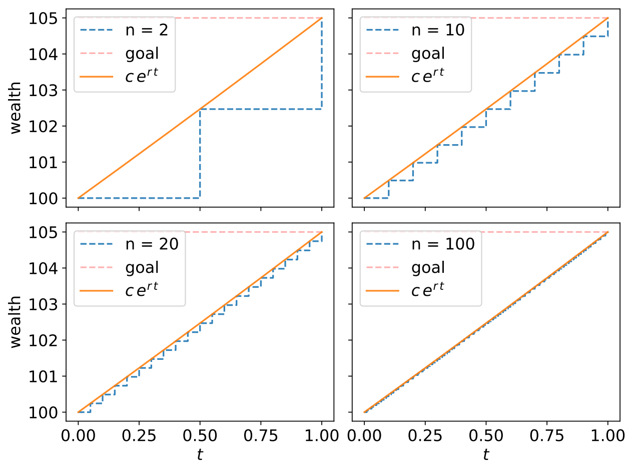

Figure 18: Wealth of an investment of £ 100 with a target 5% interest rate and various choices of compounding periods $n \in \{2, 10, 20, 100\}$.

One can show that $\lim_{n\to\infty} r^{(n)} = \log(1+R)$ – see next figure for an examle when $R = 0.05$.

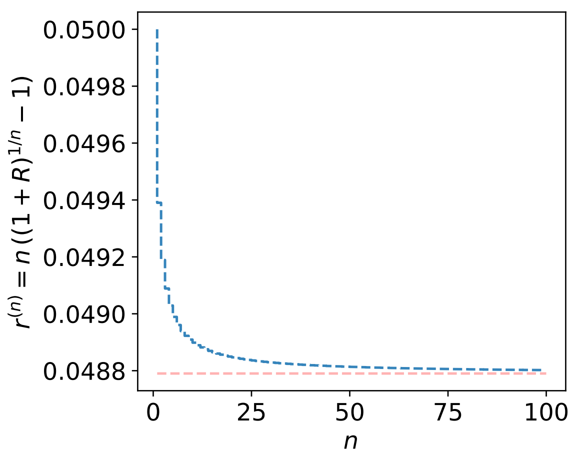

Figure 19: Function $n \to n\left((1+R)^{1/n} - 1\right)$ and its limit $\log(1+R)$.

Note that

$$e^{t\,r} = \left(1 + \frac{r^{(n)}}{n}\right)^{t\,n} \tag{208}$$

for $r = \log(1+R)$. And when n is large, an investment that compounds every $1/n$ units of time with compounded rate $r^{(n)}$ (equivalently, an effective rate $r^{(n)}/n$) produces a trajectory for the cash process that is close to that of $e^{r\,t}$ – see Figure 8.1.

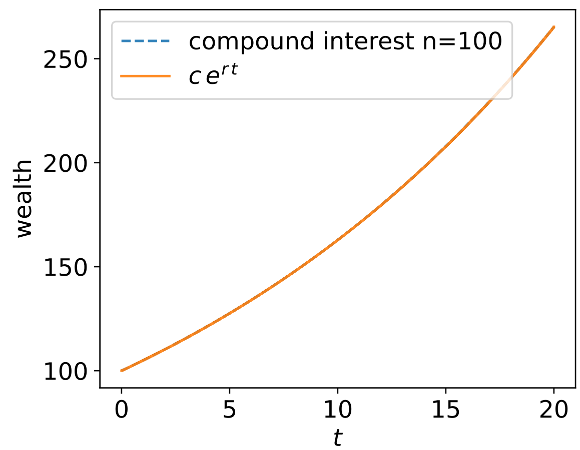

Figure 20: Continuous compounding across time $t \in [0, 20]$ and $r = \log(1 + R)$ with $R = 0.05$.

Intuitively, continuous compounding allows the investor to amend their investments in continuous time. The rate r is to be understood as the rate of return that the investor receives over an infinitesimal period of time, for example, if one invests c over the period $[t, t + \Delta)$ for some small Δ, then one receives $c(1 + r\,\Delta)$ for the investment. That is, the investor receives $c\,r\,\Delta$ in interests.

8.2 The self-financing equation

Definition 5.2 introduces the notion of self-financed strategy in equation (138). Let $(\phi_t, \psi_t)_{t\in[0,T]}$ be a trading strategy, then for $t \in [0, T]$ the quantity ϕ_t is the number of shares that the investor holds at time t and ψ_t is the number of bonds that the investor holds at time t. The value of the investor's position at time t is

$$V_t = \phi_t\,S_t + \psi_t\,B_t\,. \tag{209}$$

Let us freeze the trading strategy from t to $t + \Delta$ for a small Δ. The key idea in the self-financed equation is that the readjustment of the portfolio can be done without addition of any external funds. Mathematically, the value of the portfolio at time $t + \Delta$ is

$$V_{t+\Delta} = \phi_{t+\Delta}\,S_{t+\Delta} + \psi_{t+\Delta}\,B_{t+\Delta}\,, \tag{210}$$

but it must be equal to

$$V_{t+\Delta} = \phi_t\,S_{t+\Delta} + \psi_t\,B_{t+\Delta}\,, \tag{211}$$

too. That is, the 'value' of the portfolio before and after readjusting the strategy should be the same if we do not inject any external money to the trading strategy. From (211) and (209) we obtain that

$$V_{t+\Delta} - V_t = \phi_t\,(S_{t+\Delta} - S_t) + \psi_t\,(B_{t+\Delta} - B_t)\,, \tag{212}$$

and (212) transforms into

$$dV_t = \phi_t dS_t + \psi_t dB_t\,, \tag{213}$$

in the limit when $\Delta \to 0$.

8.3 Calculations for the semi-static hedging argument

Here I sketch the proof of Proposition 7.1.

Let $\gamma = \frac{1}{2} - r/\sigma^2$. We note that

$$\left(\frac{S_T}{S_t}\right)^{2\,\gamma} = \left(e^{\left(r-\frac{1}{2}\,\sigma^2\right)(T-t)+\sigma\,(W_T^{\mathbb{Q}}-W_t^{\mathbb{Q}})}\right)^{2\,\gamma} \tag{214}$$

$$= e^{2\,\gamma\,\left(r-\frac{1}{2}\,\sigma^2\right)(T-t)+2\,\gamma\,\sigma\,(W_T^{\mathbb{Q}}-W_t^{\mathbb{Q}})}\,. \tag{215}$$

We see that $2\,\gamma = 1 - 2\,r/\sigma^2$ and

$$\left(1-\frac{2\,r}{\sigma^2}\right)\left(r-\frac{1}{2}\,\sigma^2\right) = 2\,r - \frac{\sigma^2}{2} - \frac{2\,r^2}{\sigma^2} \tag{216}$$

$$= -\frac{1}{2}\left(\sigma^2 - 4\,r + 4\,\frac{r^2}{\sigma^2}\right) \tag{217}$$

$$= -\frac{1}{2}(2\,\gamma\,\sigma)^2\,, \tag{218}$$

and

$$\frac{S_t^2}{S_T} = S_t\,e^{-\left(r-\frac{1}{2}\,\sigma^2\right)(T-t)-\sigma\,(W_T^{\mathbb{Q}}-W_t^{\mathbb{Q}})}\,. \tag{219}$$

The above identities imply that

$$\mathbb{E}^{\mathbb{Q}}\left[\left(\frac{S_T}{S_t}\right)^{2\gamma} f\left(\frac{S_t^2}{S_T}\right)\,\middle|\,\mathcal{F}_t\right] = \mathbb{E}^{\mathbb{Q}}\left[e^{2\,\gamma\,\sigma\,(W_T^{\mathbb{Q}}-W_t^{\mathbb{Q}})-\frac{1}{2}\,(2\,\gamma\,\sigma)^2}\, f\left(S_t\,e^{-\left(r-\frac{1}{2}\,\sigma^2\right)(T-t)-\sigma\,(W_T^{\mathbb{Q}}-W_t^{\mathbb{Q}})}\right)\,\middle|\,\mathcal{F}_t\right]$$

$$= \mathbb{E}^{\tilde{\mathbb{Q}}}\left[f\left(S_t\,e^{-\left(r-\frac{1}{2}\,\sigma^2\right)(T-t)-\sigma\,(\tilde{W}_T-\tilde{W}_t)-2\,\sigma^2\,\gamma\,(T-t)}\right)\,\middle|\,\mathcal{F}_t\right] \tag{220}$$

where $\tilde{\mathbb{Q}}$ is a measure equivalent to $\mathbb{Q}$ and by Girsanov's theorem $\tilde{W}_t = W_t^{\mathbb{Q}} - 2\,\gamma\,\sigma\,t$ is a Brownian motion under $\tilde{\mathbb{Q}}$. Lastly, after some simplifications we obtain that

$$\mathbb{E}^{\mathbb{Q}}\left[\left(\frac{S_T}{S_t}\right)^{2\gamma} f\left(\frac{S_t^2}{S_T}\right)\,\middle|\,\mathcal{F}_t\right] = \mathbb{E}^{\tilde{\mathbb{Q}}}\left[f\left(S_t\,e^{\left(r-\frac{1}{2}\,\sigma^2\right)(T-t)-\sigma\,(\tilde{W}_T-\tilde{W}_t)}\right)\,\middle|\,\mathcal{F}_t\right] \tag{221}$$

and the rhs coincides with

$$e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\left[f(S_T)\,\middle|\,\mathcal{F}_t\right]\,. \tag{222}$$